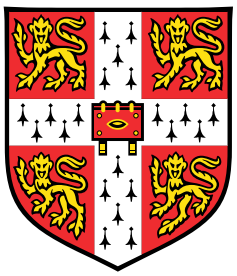
Computer-Generated Holography for Areal Additive Manufacture



Peter J. Christopher

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Peterhouse May 2021

To my wife, Bethan, for allowing me to spend more time with this than with her over the last three years. I love you!

To my mother, Sarah, for the time spent soldering my first electronics kit and sorting my Lego collection. I expect an apple pie for this!

To my father, Colin, for the hours spent trying to teach a nine-year-old pure mathematics. Remember, Dad, an engineer is "someone who applies their brain to the real world!"

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements. This dissertation contains fewer than 72,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 165 figures.

Peter J. Christopher
May 2021

Abstract

Computer-Generated Holography for Areal Additive Manufacture

Peter J. Christopher

With a market of approximately $10B, additive manufacture (AM) is an exciting next-generation technology with the promise of significant environmental and societal impact. AM promises to help reduce emissions and waste during manufacture while improving sustainability. Widely used in applications from hip implants to jet engines, AM remains the domain of experts due to the material and thermal challenges encountered.

AM in metals is dominated by Laser Powder Based Fusion (L-PBF). Powder is spread in layers 10s of microns thick and selectively melted by scanning a small laser spot heat source over the bed.

Traditional AM systems have limited ability to manage or compensate for heat generated. The rapidly moving heat source spot results in high thermal cycling and is a major influence on residual stress and distortion. Mechanical limitations in the galvoscanner mean that over- or under-heating is common and can lead to voids, boiling and spatter. The scale difference between the part size and the spot size means that predictive modelling is beyond the scope of even today's best computing clusters. These factors have led to frequent inability to ensure part quality without physical prototyping and destructive testing.

This thesis sets out initial research into creating a radically new AM process that uses computer-generated holography (CGH) to produce complex light patterns in a single pulse. Projecting power to the whole layer at once will mean that the thermal properties of the powders before and after writing can be factored into the process hologram and part design. It will also significantly reduce thermal gradients and melt-pool instability.

The fields of additive manufacture and computer-generated holography are introduced in Chapter 1. Chapters 2 and 3 then provide more detail on CGH and AM modelling respectively.

The first deliverable, a reusable software package capable of generating holograms, is presented in Chapter 4. Algorithms developed for the project are introduced in Chapter 4.3.

The first project demonstrator, an AM machine capable of printing in resins using holographic projection is discussed in Section 6.2. This shows performance comparable to modern 3D printing machines and highlights the applicability of computer-generated holography to areal processes. Section 6.3 then discusses the ongoing development of a metal powder demonstrator. As this PhD forms the first stage of a larger project, only preliminary work on the powder demonstrator is discussed. Chapter 7 then draws conclusions and outlines the way forward for future research.

The thesis appendices then discuss an in-depth discussion of algorithm performances in Appendices A and B. Appendices C and D then discuss digressions into the implement Appendices E and F present a laser induced damage threshold (LIDT) measurement system developed. Finally, Appendices G and H provide more detail on the software developed and Appendix I gives links to additional project resources.

Versions

Due to requirements on word and figure count, this thesis is available in two versions. This version is the one submitted to Cambridge University in July 2021. This thesis is self-contained and can be understood on its own.

An alternative version can be obtained from that author. This contains additional material for ease of comprehension as well as further references, appendices and explanatory figures.

Peter J. Christopher
May 2021

Acknowledgements

I would like to thank Prof Tim Wilkinson of the University of Cambridge for his academic supervision and guidance throughout this PhD. Thanks are also due to Prof Bill O'Neill of the University of Cambridge for his advice and support over the last three years.

Thanks are also due in no particular order to:

1. Ms Victoria Barrett for her invaluable help while starting the PhD, for her editing of the HoloGen application and just generally kicking ass.
2. Dr George Gordon for many a useful conversation as well as putting me onto the possibility of quantum generation algorithms.
3. Dr Calum Williams for his advice and generally putting up with my sense of humour. Also, you can finally retrieve all the lab equipment I 'borrowed'.
4. Mr Ralf Mouthaan for the many hours of discussions and helpful advice on holography, holography algorithms and their applications as well as forcing me to do experiments!
5. Mr Ralf Mouthaan, Ms Sophia Gruber and Mr Daoming Dong for their translation work on the HoloGen application.
6. Mr Daoming Dong for his assistance in referencing journal publications that are drawn on in chapters 1 and 2.
7. Mr Vamsee Bheemireddy, Mr Ralf Mouthaan, Ms Sophia Gruber, Mr Daoming Dong, Mr Fan Yang, Mr Andrew Kadis, Mr Youchao Wang, Mr Miguel El Guendy, Mr John Wheatley, Mr Xia Hua and Ms Jana Skirnewskaja for their camaraderie and support throughout the project.
8. Dr Nadeem Gabbani for his assistance in designing and building the powder demonstration system.

9. The entire Centre for Advanced Photonics and Electronics (CAPE) technical team for their help and patience with my many requests! In particular Mr Ady Ginn, Mr Stephen Drewitt, Mr Mark Barnett, Mr Joe Smith and Mr Dave Edwards.

10. My mother, Sarah Christopher, for the countless hours spent proofreading.

11. My father, Colin Christopher, for many useful discussions on the geometry of hologram generation.

12. The many collaborators on the papers published from this work including Mr Vamsee Bheemireddy, Mr Ralf Mouthaan, Ms Sophia Gruber, Mr Daoming Dong, Mr Fan Yang, Mr Andrew Kadis, Mr Youchao Wang, Mr Vamsee Bheemireddy, Mr Jamie Lake, Dr Catherine Fitzpatrick, Dr Abby Wilson, Dr Travis Sawyer, Dr Jack Alexander-Webber, Dr Calum Williams, Dr George Gordon, Dr Hannah Joyce, Dr Nicholas Alexander, Dr Barnaby Dobson, Prof Sarah Bohndiek, Prof Bill O'Neill and Prof Tim Wilkinson.

13. Last, but by no means least, my wife Bethan for supporting and putting up with me over the last three years!

Table of Contents

List of Figures

List of Tables

List of Publications

For a list of posters and presentations please contact the author.

Journal Papers, Articles and Letters - Published

1. Andrew Kadis, Youchao Wang, Daoming Dong, Ralf Mouthaan, Peter J. Christopher and Timothy D. Wilkinson, “HoloBlade: An Open-Hardware Spatial Light Modulator Driver Platform for Holographic Displays", *Applied Optics*, In Production
   DOI: 10.1364/AO.404345
2. Peter J. Christopher, Ralf Mouthaan, Miguel El Guendy and Timothy D. Wilkinson, “Linear time algorithm for phase sensitive holography", *Optical Engineering*, (2020), Volume 59, Issue 8,
   DOI: 10.1117/1.OE.59.8.085104
3. Catherine Fitzpatrick, Abby Wilson, Travis Sawyer, Peter Christopher, Timothy Wilkinson, Sarah Bohndiek, George Gordon, “Robustness to misalignment of low-cost, compact wide-field quantitative phase imaging architectures", *OSA Continuum*, (2020),
   DOI: 10.1364/OSAC.395498
4. Peter J. Christopher, Ralf Mouthaan and Timothy D. Wilkinson, “Sympathetic quantisation - a new approach to hologram quantisation", *Optics Communications*, (2020), Volume 473,
   DOI: 10.1016/j.optcom.2020.125883
5. Peter J. Christopher, Ralf Mouthaan, George S. D. Gordon and Timothy D. Wilkinson, “Holographic Predictive Search: Extending the Scope", *Optics Communications*, (2020), Volume 467,
   DOI: 10.1016/j.optcom.2020.125701
6. Peter J. Christopher and Timothy D. Wilkinson, “Relative limitations of increasing modulation levels in computer generated holography", *Optics Communications*, (2020),

Volume 462,
DOI: 10.1016/j.optcom.2020.125353

7. Peter J. Christopher, Barnaby Dobson and Nicholas A. Alexander, “Exploring the Dynamics of Base-Excited Structures Impacting a Rigid Stop.", *Mathematical Problems in Engineering* (2020) Volume 2020, ID 6721025,
DOI: 10.1155/2020/6721025
8. Youchao Wang, Daoming Dong, Peter J. Christopher, Andrew Kadis, Ralf Mouthaan, Fan Yang and Timothy D. Wilkinson “Hardware Implementations of Computer Generated Holography: A Review", (2020) *Optical Engineering*, Volume 59, Issue 10, pp. 1–30
DOI: 10.1117/1.OE.59.10.102413
9. Peter J. Christopher, Ralf Mouthaan, Vamsee Bheemireddy, and Timothy D. Wilkinson, “Improving performance of single-pass real-time holographic projection", *Optics Communications* (2020), Volume 457, pp. 1456-1462,
DOI: 10.1016/j.optcom.2019.124666
10. Peter J. Christopher, Youchao Wang, and Timothy D. Wilkinson, “Predictive search algorithm for phase holography", *Journal of the Optical Society of America A* (2019), Volume 36, Issue 12, pp. 2068-2075,
DOI: 10.1364/JOSAA.36.002068
11. Peter J. Christopher, Jamie Lake, Daoming Dong, Hannah Joyce, and Timothy D. Wilkinson, “Improving holographic search algorithms using sorted pixel selection", (2019), *Journal of the Optical Society of America A*, Volume 36, Issue 9, pp. 1456-1462,
DOI: 10.1364/JOSAA.36.001456

Journal Papers, Articles and Letters - Under Review

12. Peter J. Christopher, Andrew Kadis, George S. D. Gordon and Timothy D. Wilkinson, “HoloGen: An open source toolbox for high-speed hologram generation", *Computational Physics Communications*, Under Review

In Proceedings - Published

13. Youchao Wang, Daoming Dong, Andrew C. Kadis, Peter J. Christopher and Timothy D. Wilkinson, “Computer-generated holography using a digital signal processor", *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019,
DOI: 10.1109/GlobalSIP45357.2019.8969314

14. Daoming Dong, Youchao Wang, Andrew C. Kadis, Peter J. Christopher and Timothy D. Wilkinson, “Fixed-Point Accuracy Analysis of 2D FFT for the Creation of Computer Generated Holograms", *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019,
DOI: 10.1109/GlobalSIP45357.2019.8969134
15. Peter J. Christopher, Ralf Mouthaan, Andrew Kadis, Nadeem Gabbani, William O’Neill, Timothy D. Wilkinson “Thermal Compensation for High Load Spatial Light Modulators in Real-Time", *CLEO*, 2020, DOI: 10.1364/10.1364/CLEO_AT.2020.ATu4K.5
16. Andrew Kadis, Daoming Dong, Youchao Wang, Peter J. Christopher, Ralf Mouthaan and Timothy D. Wilkinson “HoloBlade: An Open Platform for Holography", *OSA Imaging and Applied Optics*, 2020, DOI: 10.1364/DH.2020.HF4D.4

In Proceedings - Accepted

17. Peter J. Christopher, Ralf Mouthaan, Andrew Kadis, Nadeem Gabbani, William O’Neill, Timothy D. Wilkinson, “Holographic additive manufacture for stereolithographic 3D printing", *Photonics West*, 2021,

Journal Papers, Articles and Letters - Preprint

18. Peter J. Christopher and Timothy D. Wilkinson, “Structure and Design of HoloGen", 2020, arXiv: 2006.10509
19. Peter J. Christopher, George S. D. Gordon and Timothy D. Wilkinson “Benchmarking the Gerchberg-Saxton Algorithm", 2020, arXiv: 2005.08623
20. Peter J. Christopher, Ralf Mouthaan and Timothy D. Wilkinson “Lookup tables for phase randomisation in hardware generated holograms", 2020, arXiv: 2004.04049
21. Peter J. Christopher, Nadeem Gabbani, William O’Neill and Timothy D. Wilkinson, “Automated laser induced damage threshold testing applied to a ferroelectric spatial light modulator", 2020, arXiv: 2004.00712
22. Peter J. Christopher and Timothy D. Wilkinson, “Variance and Error in One-Step Phase-Retrieval", 2019, arXiv: 1911.00045
23. Peter J. Christopher, Ralf Mouthaan, John P. Freeman and Timothy D. Wilkinson, “Improving pixel differentiation in holographic images", 2019, arXiv: 1912.12196

Media

24. Peter J. Christopher interviewed by Nadeem Gabbani, “Phenomics: A Medical Revolution", *Naked Scientists*, BBC Radio 5, November 15th 2019, 21:04 - 26:10

25. Peter J. Christopher interviewed by Madhav Malhotra, “Holography: What’s Next After VR/AR", *The Knowledge Archives*, August 14th 2020,

Software Packages

26. Peter J. Christopher, "HoloGen - Computer Generated Holography Suite", *www.gitlab.com/CMMPEOpenAccess/HoloGen*, 2018

Grants and Awards

27. Peter J. Christopher and Timothy D. Wilkinson, “Additive manufacturing using holographic beam shaping”, *Engineering and Physical Sciences Research Council*, 2020, £362,362
28. Peter J. Christopher, Jamie Lake, Timothy D. Wilkinson and Hannah Joyce, “Eye tracking for 3D holographic displays", *Observatory for Human Machine Collaboration (OHMC)*, 2020, £15,768
29. Jamie Lake, Peter J. Christopher, Ralf Mouthaan, Timothy D. Wilkinson and Hannah Joyce, “Focal depth tracking for improved three-dimensional holographic displays", *Centre for Advanced Photonics and Electronics (CAPE) Acorn Post-graduate Research Award*, 2020, £2,000
30. Daoming Dong, Youchao Wang, Andrew C. Kadis, Peter J. Christopher and Timothy D. Wilkinson, “Interfacing a high speed ferroelectric spatial light modulator", *Centre for Advanced Photonics and Electronics (CAPE) Acorn Post-graduate Research Award*, 2019, £3,000
31. Andrew C. Kadis, Youchao Wang, Daoming Dong, Peter J. Christopher, Ralf Mouthaan and Timothy D. Wilkinson, “Low cost SLM interface board for advanced microscopy and holographic display of biological structures”, *OpenPlant Biomaker Challenge*, 2019, £2,750

Lectures

32. Peter J. Christopher, "Fantastic holograms and how to generate them", *4B11: Photonic Systems*, (2019)

Nomenclature

Roman Symbols

a_n	Weighting Coefficient of the n'th Zernike Polynomial	
b	Depth of Focus or Confocal Parameter	m
P	Magnetic Field	T
c	Speed of Light in a Vacuum $= 2.9979 \times 10^8 \mathrm{m/s}$	$\mathrm{m/s}$
c_1	Structural Similarity Index Stabilisation Factor 1	
c_2	Structural Similarity Index Stabilisation Factor 2	
C_p	Specific Heat Capacity	$\mathrm{J/(m^3 \cdot K)}$
D	Width of Deposit	m
d	Distance between lens and SLM	m
D_0	Focal Point Diameter	m
D^E	Elastic Stiffness Matrix	$\mathrm{N/m}$
D^{EP}	Elastic-Plastic Stiffness Matrix	$\mathrm{N/m}$
E	Elastic / Young's Modulus	$\mathrm{kg \cdot m/s^2}$
E	Electric Field	$\mathrm{V/m}$
E	Error / Simulated Annealing Cost Function	
e	Ellipticity	
E_{max}	Major Axis of the Polarization Ellipse	
E_{min}	Minor Axis of the Polarization Ellipse	
E_p	Pulse Energy	J
E_ψ	Polarisation Amplitude Component at Angle ψ	

E_v	Volumetric Heat Input	$\mathrm{J/m^3}$
f	Distribution Factor	
f	Focal Length	m
F	Mask Helper Field	
F	Laser Fluence	$\mathrm{J/m^2}$
f_n	Height of Peak or Valley	m
G	Thermal Gradient	$\mathrm{K/m}$
g	Acceleration due to gravity $= 9.81\mathrm{m \cdot s/s}$	$\mathrm{m \cdot s/s}$
G	Ideal Hologram	
$G^m(\theta)$	Zernike Angular Polynomial	
H	Hadamard Matrix	
h	Latent / Sensible Heat	$\mathrm{J/K}$
H'	Quantized Hologram	
H_{quant}	Quantized Hologram Function	
H_{uncor}	Hologram Function Prior to Aberration Correction	
H	Hologram	
H_n	Sub-frame n of hologram H	
h_c	Convective Heat Transfer Coefficient	$\mathrm{W/(m^2 \cdot K)}$
H_{cor}	Hologram Function After aberration Correction	
I	Illumination	$\mathrm{W/m^2}$
I_0	Peak Beam Intensity	$\mathrm{W/m^2}$
I	Moment of Inertia	$\mathrm{kg/m^2}$
J	Mean Squared Error	
J	Jones Matrix	
J_i	Bessel Function Number i	
k	Propagation Constant	
k	Thermal Conductivity	$\mathrm{W/(m \cdot K)}$
k	Wavenumber	m
k_1	Structural Similarity Index Constant	

k_2	Structural Similarity Index Constant	
K	Mask Helper Field	
c_L	Pixel Dynamic Range	
L	Characteristic Length	m
L	Mask Helper Field	
M	Mask Function	J/m^2
M^2	Beam Quality Factor	
M_a	Marangoni Number	
M_n	Molecular Weight of Element n	kg/mol
N	Number of OSPR Sub-Frames	
n	Refractive Index	
n_0	Refractive Index in Air	
$\widetilde{n}$	Complex Refractive Index	
n_g	Refractive Index in Gas	
N_n^m	Zernike Normalisation Constant	
$N(\mu, \sigma^2)$	Normal Distribution with Mean μ and Variance σ^2	
N_u	Width of Replay Field in Pixels	
N_v	Height of Replay Field in Pixels	
N_x	Width of Diffraction Field in Pixels	
N_y	Height of Diffraction Field in Pixels	
P	Heat Source Power	W
P	Simulated Annealing Probability Function	
P_e	Peclet Number	
Φ	Hologram Aberration Phase Profile	
P_n	Equilibrium Vapor Pressure of Element n	Pa
$Q*$	Non-Dimensional Heat Input	
r	Radial Distance from Heat Source Axis	m
R'	Constrained Replay Field	
R	Replay Field	

T	Target Image	
R_a	Average Surface Roughness	m
R_p	Reflectance of p-Polarised Light	
R_s	Reflectance of s-Polarised Light	
S_j	Momentum Equation Source Term	
S_n	Stokes Parameter n	
t	Time	s
t_0	Initial Temperature	K
t	Simulated Annealing Temperature	K
t_{coeff}	Simulated Annealing Temperature Coefficient	
T_a	Ambient Temperature	K
t_l	Layer Thickness	m
T_p	Peak Temperature	K
T_p	Transmissivity of p-Polarised Light	
T_s	Transmissivity of s-Polarised Light	
u	First Spatial Frequency	$1/\mathrm{m}$
v	Second Spatial Frequency	$\mathrm{cycles}/\mathrm{m}$
$V_n^m(\rho)$	Zernike Radial Polynomial	
v	Speed of Light in a Medium	m/s
V_{in}	SLM Input Amplitude	m
V_{out}	SLM Output Energy	J
w_0	Beam Waist	m
w_b	Heat Source Radius	m
w_j	Zernike Polynomial Weighting	
x	First Diffraction Plane Axis	m
y	Second Diffraction Plane Axis	m
z_R	Rayleigh Distance	m

Greek Symbols

α	Thermal Diffusivity	m^2/s

β	Quantization Scale Factor	
t	Fraction of Destroyed Pixels	
χ	Ellipticity Angle	
Δ	Pixel Pitch	m
ΔH	Latent Heat	J/K
ΔT	Temperature Difference	K
ΔT_{tot}	Total Undercooling	K
$d\varepsilon$	Total Strain Increment	
$d\varepsilon^E$	Total Elastic Strain Increment	
$d\varepsilon^P$	Total Plastic Strain Increment	
$d\varepsilon^T$	Total Thermal Strain Increment	
$d\varepsilon^P$	Total Volumetric Strain Increment	
δ_0	Kronecker Delta Function	
ε_0	Permittivity of Free Space	F/m
ε	Error	%
ε	Strain	
ε	Eccentricity	
ε	Emissivity	W/m^2
ε_C	Cooling Rate	K/s
ε_e	Elastic Strain	
$\varepsilon_{e,max}$	Maximum Elastic Strain	
ε_o	Inelastic Strain due to Creep	
ε_p	Plastic Strain	
$\varepsilon_\parallel$	Parallel Dielectric Constant	
$\varepsilon_\perp$	Perpendicular Dielectric Constant	
η	Fraction of on Pixels	
η	Relative Phase Retardation	°
η	Second Replay Field Axis	m
η_l	Deposit Layer Absorption Coefficient	

Γ	Gamma Function	
γ	Display Gamma Factor	
γ	Surface Tension	$\mathrm{N/m}$
Γ	Wave Plate Retardation	rad
Δ_λ	Spectral Bandwidth	m
$\Delta_{\lambda,lim}$	Limiting Spectral Bandwidth	m
λ	Wavelength	m
λ	Laser Absorptivity	
λ_c	Positive Fraction due to Vaporised Atom Condensation	
μ	Distribution Mean	
μ_0	Permeability of Free Space	$\mathrm{H/m}$
μ	Dynamic Viscosity	$\mathrm{N \cdot s/m^2}$
μ_R	Mean of Replay Image, R	
μ_T	Mean of Target Image, T	
Ω	Calculation Region	
ϕ	Phase	rad
ψ	Angle between Solidification Interface and Normal Direction	rad
ψ	Orientation Angle	
ρ	Radial Component of the Zernike Unit Circle	
σ	Stress	$\mathrm{N/m^2}$
σ_R^2	Variance of replay image, R	
σ_{SB}	Stefan-Boltzmann Constant $= 5.67 \times 10^{-8} \mathrm{W/(m^2 \cdot K)}$	$\mathrm{W/(m^2 \cdot K)}$
σ^2	Variance	
σ_T^2	Variance of Target Image, T	
σ_{TR}	Covariance of Target Image, T and Replay Image, R	
τ	Characteristic Time Scale	
τ_M	Marangoni Stress	$\mathrm{N/m^2}$
Θ_{rec}	Recreated Phase	
Θ	Phase	rad

θ	Angular Component of the Zernike Unit Circle	
θ	Liquid Crystal Switching Angle	rad
θ	SLM Viewing Angle	rad
θ_{chrom}	Chromatic Spread	m
θ_i	Incidence Angle	rad
$\theta_{diffmax}$	Maximum Diffraction Angle	rad
θ_{ps}	Phase Spread	m
θ_r	Reflectance Angle	rad
Θ_{rand}	Randomised Phase	rad
θ_{res}	Limiting Replay Field Resolution	m
θ_{src}	Spread due to Light Source Diameter	m
θ_t	Transmission Angle	rad
Υ	Attenuation Function due to Pixel Shape	
ω	Frequency	/m
ξ	First Replay Field Axis	m

Subscripts

j	Zernike Polynomial Noll Index
m	Azimuthal Frequency of a Zernike Polynomial Element
n	Order of a Zernike Polynomial Element

Other Symbols

$\mathbb{C}$	The Set of Complex Numbers
$\chi^2(k)$	Chi-Squared Distribution
$E[\mathbb{X}]$	Expected Value of Random Variable $\mathbb{X}$
$\mathscr{F}$	Fourier Transform
$\mathscr{F}^{-1}$	Inverse Fourier Transform
N	Normal Distribution
$\mathscr{FR}$	Fresnel Transform
$\mathscr{I}$	Imaginary Part
$\mathscr{L}$	Laplace Transform

$\mathbb{N}$	The Set of Natural Values
$N(\mu, \sigma^2)$	Normal Distribution
$O()$	Big 'O' Notation for Algorithm Performance
p	Probability Density Function
$\mathbb{Q}$	The Set of Rational Values
$\mathscr{R}$	Real Part
$R(\nu, \sigma)$	Rice Distribution
$R(\sigma)$	Rayleigh Distribution
$V[\mathbb{X}]$	Variance of Random Variable $\mathbb{X}$
$\mathbb{R}$	The Set of Real Numbers
$\mathbb{Z}$	The Set of Integer Numbers
$\bigotimes$	Tensor Product
$a \angle b$	Increment the phase of s by b
$\cdot$	Dot or Hadamard (Point-wise) Product
$*$	Convolution
$\circ$	Function Multiplication
∇	Three Dimensional Gradient Operator
$\times$	Cross or Cartesian Product

Acronyms / Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
3DP	3D Printing
AC	Alternating Current
Ad-OSPR	Adaptive One-Step Phase Retrieval Algorithm
AFM	Atomic Force Microscopy
ALE	Arbitrary Lagrangian-Eulerian
AM	Additive Manufacture
ANN	Artificial Neural Network
API	Application Programming Interface

AS	Acoustic Spectroscopy
ASTM	American Society for Testing and Materials
B2F	Buy-to-Fly Ratio
BIS	Beam Interference Solidification
BJ	Binder Jetting
BQF	Beam Quality Factor
BS	Beam Splitter
CDS	Concurrent Direct Search
CSA	Concurrent Simulated Annealing
CAD	Computer Aided Design
CAM	Computer Aided Manufacture
CAN	Controller Area Network
CCD	Charge Coupled Device
CGH	Computer-Generated Hologram
CLT	Central Limit Theorem
CLT	Concurrent Linear Time
CMOS	Complimentary Metal-Oxide Semiconductor
CMT	Cold Metal Transfer
CNN	Convolutional Neural Network
COM	Component Object Model
CPU	Computer Processing Unit
CSI	Coherence Scanning Interferometry
CT	Computer Tomography
CTE	Coefficient of Thermal Expansion
CTG	Cooling Temperature Gradient
CUDA	Compute Unified Device Architecture
cuFFT	Cuda Fast Fourier Transform Library
CW	Continuous Wave
CWO	Computational Wave Optics

D2M	Direct to Metal
DBN	Deep Belief Network
DBS	Direct Binary Search
DC	Direct Current
DCT	Discrete Cosine Transform
DED	Directed Energy Deposition
DF	Diffraction Field
DFT	Discrete Fourier Transform
DH	Digital Holography
DHM	Digital Holographic Microscopy
DLL	Dynamic Link Library
DMD	Digital Micro-Mirror Device
DMLS	Direct Metal Laser Sintering
DoF	Degrees of Freedom
DP	DisplayPort
DS	Direct Show
DS	Direct Search
DSP	Digital Signal Processor
DWT	Discrete Wavelet Transform
EB-DED	Electron Beam Directed Energy Deposition
EB	Electron- / E-Beam
EBM	Electron Beam Melting
EB-PBF	Electron Beam Powder Bed Fusion
ED	Error Diffusion
EDS	Energy Dispersive X-Ray Spectroscopy
EM	Electromagnetism
EP	Error Propagation
EULA	End User License Agreement
FBS	Fundamental Beam-mode Shaping

FC	Faraday Cup
FDM	Finite Difference Method
FDM	Fused Deposition Modelling
FEA	Finite Element Analysis
FF	Far-Field
FFT	Fast Fourier Transform
FICH	Fresnel Incoherent Correlation Holography
FLC	Ferroelectric Liquid Crystal
FLCOS	Ferroelectric Liquid Crystal on Silicon
FL	Focal Length
FoV	Field of View
FPGA	Field Programmable Gate Array
FPS	Frames per Second
FPU	Floating Point Unit
FSD	Focal Spot Diameter
FT	Fourier Transform
FVM	Focus Variation Microscopy
FWHM	Full-Width Half-Maximum
GA	Gas Atomisation
GIC	Global Iterative Coding
GLV	Grating Light Valve
GMa-DED	Gas Metal Arc Directed Energy Deposition
GMA	Gas Metal Arc
GPIB	General Purpose Interface Bus
GPL	General Public License
GPU	Graphics Processing Unit
GS	Gerchberg-Saxton
GUI	Graphical User Interface
HDMI	High Definition Multimedia Interface

HDP	Holographic Display Panel
HeNe	Helium Neon
HI	Holographic Interferometry
HIP	Hot Isostatic Pressing
HIS	Holographic Interference Solidification
HMBIL	Holographic Multi-Beam Interference Lithography
HPS	Holographic Predictive Search
HSA	Holographic Search Algorithm
HSLA	Holographic Stereolithography
i.i.d.	Independent indentically distributed
I^2C	Inter-Integrated Circuit
IDE	Integrated Development Environment
IFFT	Inverse Fast Fourier Transform
IFTA	Iterative Fourier Transform Algorithm
IO	Input/Output
IPA	Isopropanol
IP	Intellectual Property
IR	Infrared
ITO	Indium Tin Oxide
JSON	JavaScript Object Notation
LAE	Large Area Electronics
LBM	Lattice Boltzmann Method
LBO	Light Blue Optics
LCD	Liquid Crystal Display
LC	Liquid Crystal
LCOS	Liquid crystal on Silicon
L-DED	Laser Directed Energy Deposition
LED	Light Emitting Diode
LENS	Laser Engineered Net Shaping

LGPL	Lesser General Public License
LIBS	Laser-Induced Breakdown Spectroscopy
LIDT	Laser-Induced Damage Threshold
LoC	Lines of Code
LOM	layered Object Manufacturing
L-PBF	Laser Powder Bed Fusion
LPD	Laser Powder Deposition
LSM	Level Set Method
LT-DS	Linear Time Direct Search
LT-HPS	Linear Time Holographic Predictive Search
LT-SA	Linear Time Simulated Annealing
LT	Liu-Taghizadeh Algorithm
LTP	Liquid Thermal Polymerization
LUT	Look-Up Table
MBIL	Multi-Beam Interference Lithography
MC	Marangoni Convection
ME	Material Extrusion
MEMS	Micro-Electromechanical System
MIG	Metal Inert Gas
MI	Multiple Inheritance
MIR	Mid-Infrared
MJ	Material Jetting
MKL	Intel Math Kernel Library
ML	Machine Learning
MOEM	Micro-Opto-Electromechanical System
MR	Mixed Reality
MSE	Mean Squared Error
NMSE	Normalised Mean Squared Error
MVC	Model-View-Controller Application Architecture

MVVMA	Model-View-ViewModel-Algorithm Application Architecture
MVVM	Model-View-ViewModel Application Architecture
NA	Numerical Aperture
ND	Neutron Diffraction
ND	Neutral Density
NF	Near-Field
NI	National Instruments
NIR	Near Infrared
NN	Nearest-Neighbour
NN	Neural Network
NNQ	Nearest Neighbour Quantisation
NNS	Near-Net Shape
NS	Net Shape
NVRTC	Nvidia Runtime Compilation
OCR	Optical Character Recognition
OE	Optical Element
OOP	Object Orientated Programming
ORA	Optimal Rotation Angle
OSH	Optical Scanning Holography
OS	Operating System
OSPR	One-Step Phase Retrieval Algorithm
PA	Plasma Arc
PBF	Powder Bed Fusion
PBS	Polarizing Beam Splitter
PDD	Power Density Distribution
PDF	Probability Density Function
PE	Polarisation Ellipse
PI	Phase Insenstive
PLS	Point Light Source

PMZ	Partially Melted Zone
PREP	Plasma Rotating Electrode Process
PRNG	Pseudo-Random Number Generator
PR	Phase Randomisation
PR	Phase Retrieval
PSA	Predictive Simulated Annealing
PSD	Particle Size Distribution
PSD	Power Spectral Density
PSI	Phase Scanning Interferometry
PSNR	Peak Signal-to-Noise Ratio
PS	Phase Sensitive
PWPS	Pixel-to-Wrapped Phase Summation Algorithm
QED	Quod Erat Demonstrandum
QPM	Quantitative Phase Microscopy
QuML	Quantisation using Machine Learning
RAII	Resource Acquisition is Initialization
RAM	Random Access Memory
RA	Rotary Atomisation
RF	Replay Field
RMS	Root Mean Square
RNG	Random Number Generator
RoI	Region of Interest
RPF	Replay Field
RP	Rapid Prototyping
RS	Ramen Spectrocopy
SA	Simulated Annealing
SB	Spectral Bandwidth
SDK	Software Development Kit
SDP	Semidefinite Programming

SEM	Scanning Electron Microscope
SfM	Structure-from-Motion
SGC	Solid Ground Curing
SGR	Solidification Growth Rate
SHS	Selective Heat Sintering
SLA	Stereolithography
SLC	Selective Laser Cladding
SLM	Selective Laser Melting
SLM	Spatial Light Modulator
SL	Sheet Lamination
SLS	Selective Laser Sintering
SM	Single Mode
SM	Subtractive Manufacture
SNR	Signal-to-Noise Ratio
SoC	System on Chip
SP	Shot Peening
SPS	Sorted Pixel Selection
SQ	Sympathetic Quantisation
SSD	Solid State Drive
SSIM	Structural Similarity Index
SSQ	Soft Sympathetic Quantisation
STL	Standard Template Library
STL	Stereolithography
STTM	Single-Transform Time-Multiplexed
TDP	Thermal Design Power
TEC	Thermo-Electric Cooler
TEM	Transmission Electromagnetic Microscopy
TEM	Transverse Electromagnetic Modes
TIG	Tungsten Inert Gas

TMA	Time-Multiplexed Algorithm
TN	Twisted Nematic
TSPR	Two-stage Sparse Phase Retrieval
UAM	Ultrasonic Additive Manufacture
UI	User Interface
USB	Universal Serial Bus
UV	Ultra Violet
UX	User Experience
VGA	Video Graphics Array
VoF	Volume of Fluid
VR	Virtual Reality
VSI	Vertical Scanning Interferometry
WAAM	Wire and Arc Additive Manufacture
WA	Water Atomisation
WCF	Windows Communication Framework
WDM	Wavelength Division Multiplexing
WD	Working Distance
WF	Wirtinger Flow
WGS	Weighted Gerchberg-Saxton
WINN	Quantum Inspired Neural Network
WLI	White Light Interferometry
WPF	Windows Presentation Framework
WRP	Wavefront Recording Plane
w.r.t.	with respect to
WSL	Windows Service Layer
WSM	Windows Service Model
WWS	Windows Web Services
XAML	eXtensible Application Markup Language
XCT	X-ray Coherence Tomography

XML	eXtensible Markup Language
XPS	X-ray Photoelectron Spectroscopy

Note on Terminology

A number of conventions are followed in this document.

- References to programming constructs such as classes and functions are shown in a `unique` font.
- References to libraries and packages are shown in *italics*.
- Complex conjugation is denoted by the use of overbar as in $\overline{x}$.
- Random variables are referred to using a blackboard font as in $\mathbb{X}$.
- Operations with real-valued inputs are shown using standard function notation while operations with discrete inputs are shown using an index format.
- The derivative of function $f(x)$ is sometimes shown as f'. Differentiation with respect to time is sometimes shown as $\dot{f}$.
- Normalised variables are shown with a circumflex as in $\hat{x}$.

While the majority of symbology used is well known, the following are less well known. The symbols '$\forall$', '$\therefore$', '$\because$' and '$\square$' are interpreted as "for all", "therefore", "because" and "QED" respectively.

Notation has been chosen to follow the best-known works on the area except when necessary to reconcile the often contradictory terminologies of light-matter interactions, additive manufacture, computer modelling and holography.

Chapter 1

Background and Introduction

Chapter 1

Background and Introduction

1.1 Introduction

This chapter provides a background for additive manufacture in Section 1.2 and holography in Section 1.3. Each field is presented in historical context and a general overview given before relevant areas are expanded on in detail. Section 1.5 then discusses previous syntheses of the two areas.

1.2 Additive Manufacture

1.2.1 Introduction

Additive manufacture (AM), 3D printing (3DP) and rapid prototyping (RP) are similar terms used to describe the production of objects by combining material in contrast to traditional subtractive processes[1]. While these terms are often conflated, typically RP refers to prototyping done during design with a focus on turnaround; AM refers to manufacture based processes where cost dominates over speed and part quality/reliability is paramount and 3DP stands between these and refers to lower volume manufacture in resins, waxes, plastics and similar. This thesis is concerned primarily with AM, both in resins, plastics and metals with a concentration on increasing production speeds using novel heat sources.

Held back for decades by the high cost of lasers and powder, hardware limitations for the computation required and an insufficient understanding of the processes involved recent years have seen an enormous uptick in the use of additive with AM manufactured parts being increasingly used in high-value applications including the aerospace, racing, sports and medical industries. Additive parts differ from their traditionally manufactured counterparts with different metallurgy, residual stresses, anisotropy, defects, thermal creep and fatigue. Handling these challenges is the primary focus of the recent renewal of interest in the topic. While AM still faces a number of challenges, it offers exciting opportunities to produce lighter, stronger, more organic geometries including lattices and shells not previously possible.

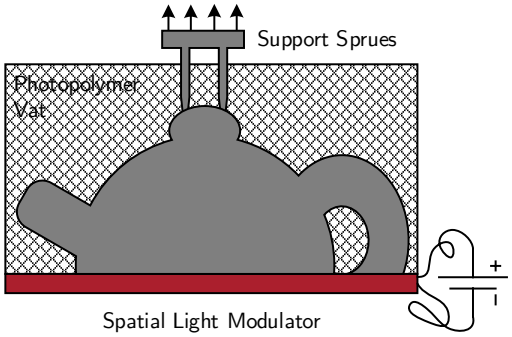
1.2.2 Processes

The American Society for Testing and Materials (ASTM) standards define two categories of AM processes: directed energy deposition (DED) and powder bed fusion (PBF)[2]. These come in a wide variety of variants summarised in Figure 1.1. Additive processes can also be categorised as direct to metal (D2M) where post-processing is limited and near-net shape (NNS) where significant post-machining is required. While detailed knowledge of the individual processes is not required for this thesis, an overview of the options along with their limitations is useful when considering alternative approaches.

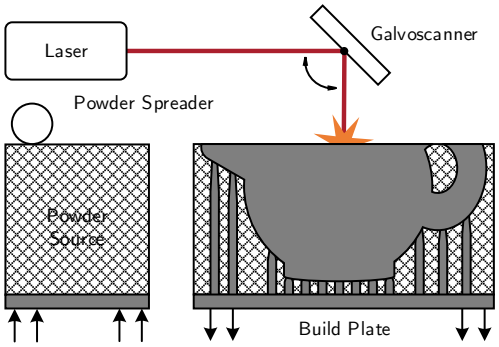
1.2.3 Powder Bed Fusion

Powder bed fusion (PBF) is dominated by two main processes laser-PBF (L-PBF) and electron beam-PBF (EB-PBF) and is perhaps the main method for AM in metals today[3]. Many variants exist including direct metal laser sintering (DMLS), electron beam melting (EBM), selective laser sintering (SLS), selective heat sintering (SHS) and selective laser melting (SLM)[4]. One fundamental difference between these is whether they use a laser power source as in Figure 1.1i or an electron source as in Figure 1.1j. A secondary distinction is whether the material is melted or sintered, *sintering* being where the material coalesces below the melting temperature without ever actually changing phase.

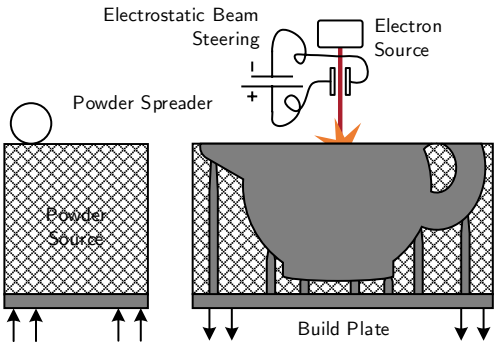
A powder spreader lays down a layer of powder, typically 10s of µms thick. A laser or electron beam heat source is then rastered over the surface using a galvoscanner in the regions where



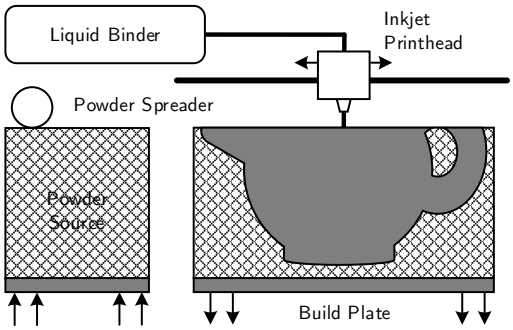
(a) Photopolymerisation



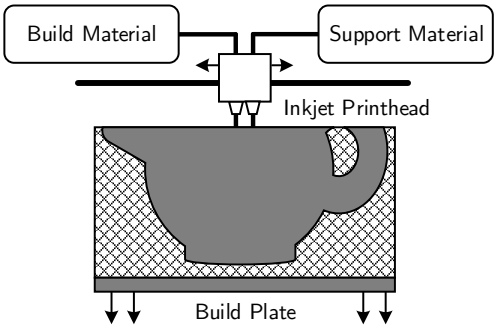
(b) Laser PBF



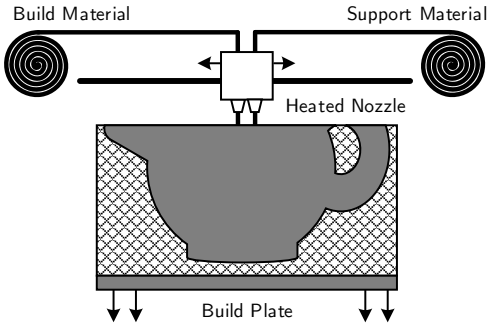
(c) Electron Beam PBF



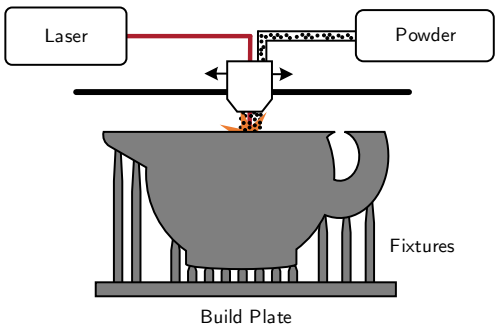
(d) Binder Jetting



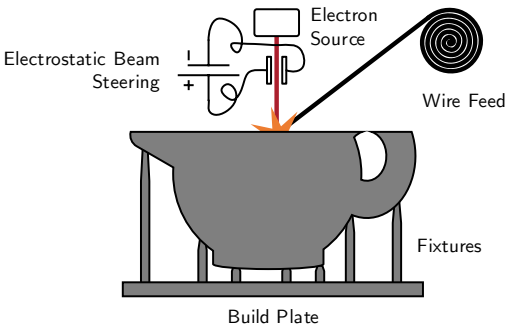
(e) Material Jetting



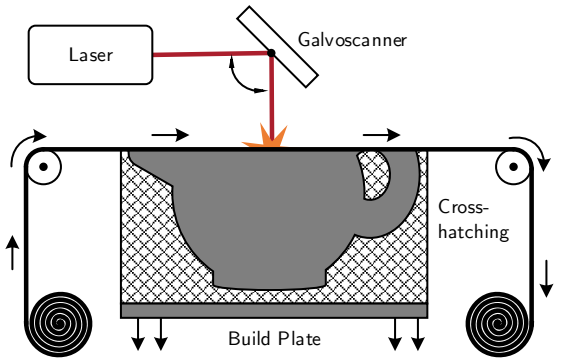
(f) Material Extrusion



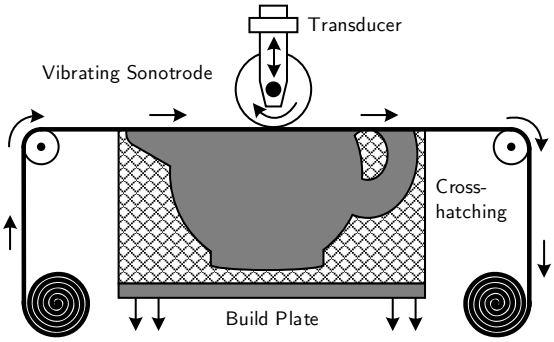
(g) Laser DED



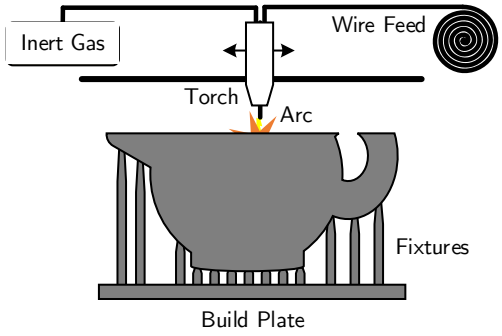
(h) Electron Beam DED



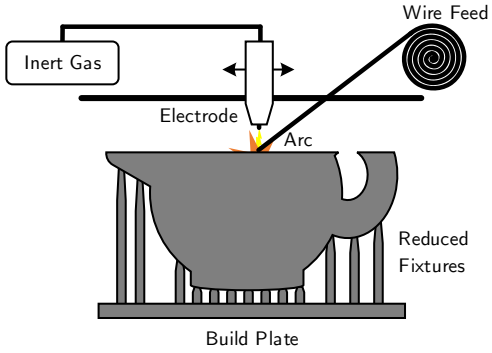
(i) Sheet Lamination



(j) Ultrasonic Additive Manufacture



(k) Gas Metal Additive



(l) Wire and Arc Additive Manufacture

Fig. 1.1 Additive manufacture processes

solidification is desired. This is repeated, building up the part in many layers. Notable in this process is that the high thermal gradients when working with metals necessitate significant support structures or fixtures to keep the part in place. While L-PBF operates in an inert environment, EB-PBF must operate in a vacuum.

The lack of moving parts means that EB-PBF can achieve scanning speeds of 100s of m/s in comparison to L-PBF's 1s of m/s. The low scanning speeds and energy input of L-PBF mean that slight inconsistencies in motion and powder can significantly change the heat input at a point, leading to boiling, particulate ejecta and deformation. EB-PBF, on the other hand, is the process of choice for producing production quality metal parts at the present moment but charging effects limit the range of available materials[5].

This report aims to introduce an alternative form of L-PBF to work on metallic parts while mitigating against the prevalent thermal issues. Variations on PBF using a blown powder stream include laser powder deposition (LPD), selective laser cladding (SLC), laser engineered net shaping (LENS).

1.2.4 Laser and Electron Beam Directed Energy Deposition

Directed energy deposition (DED) is similar to PBF and also comes in laser (L-DED) and electron (EB-DED) variants shown in Figures 1.1g and 1.1h respectively. Material sources can differ widely with wire and powder dominating the market at the time of writing. As with PBF, significant fixturing is required for metallic parts. Powder wastage and low part quality means that this process has a small market share next to comparable PBF systems[3].

1.2.5 Gas Metal Additive

Gas metal additive (GMA) is similar to the DED methods discussed above except that it uses an arc-based heat source[6]. Figure 1.1k and Figure 1.1l show the two dominant processes. The first shows a system based on a cold metal transfer (CMT) process similar to that used in a metal inert gas (MIG) welder. The CMT process uses detailed computer control to reduce the amount of melting required[7]. Instead, a cold welding process is observed similar to that encountered in spacecraft[8]. The wire is used as a perishable electrode with the part acting as a ground for the arc.

Similar to GMA is wire and arc additive manufacture (WAAM) shown in Figure 1.1l. This uses a tungsten inert gas (TIG) based process with a non-perishable electrode with wire added externally. This offers the greater precision of TIG welding but reduces the processing speed. Both processes are increasingly being used in industry for applications including panel manufacture where they can cut the buy-to-fly (B2F) ratio from around 100 in some aerospace applications to around 2 after post-machining[9].

1.2.6 Extrusion and Jetting

Three similar processes dominate the plastics AM market alongside L-PBF: material extrusion (ME), material jetting (MJ) and binder jetting (BJ). Binder jetting, Figure 1.1d, works on a powder bed layer-wise with a nozzle extruding binding agent in the desired regions. As with most plastics based operations, the requirements for fixturing are much lower and the part can often be supported only by the surrounding powder. Material jetting, Figure 1.1e, drops the powder bed and instead uses a liquid or pair of liquids which cure when exposed to light to form the solid regions. Material extrusion, Figure 1.1f, works by extruding a wire of

a, typically thermoplastic, material that can be melted by a heated nozzle. This last process constitutes the majority of the home 3DP market.

1.2.7 Photopolymerisation

Figure 1.1a shows one of the conceptually simplest 3DP processes, *photopolymerisation*. A Liquid Crystal Display (LCD) is placed underneath a vat of curable liquid resin and the displayed light pattern changed in sync with the part being drawn out of the liquid. Photopolymerisation is currently the fastest printing methods on the market. The choice of SLM is of great interest and a step-change in the power capacity has the potential to revolutionise the industry[10]. Variations on this process include Stereolithography (SLA), solid ground curing (SGC), liquid thermal polymerization (LTP), beam interference solidification (BIS), holographic interference solidification (HIS)[3].

1.2.8 Laminated Object Manufacturing

Two final systems deserve mention: sheet lamination (SL) and ultrasonic additive manufacture (UAM) shown in Figure 1.1i and Figure 1.1j. Both of these fall under the category of laminated object manufacturing (LOM) and operate by affixing layers of existing material to a layered substrate using either heat - SL - or pressure - UAM[11].

The choice of process is decided primarily by the part quality required as well material and finished weight. For mass producing macro-scale production-quality parts with limited required post-processing, PBF is currently the only system with the promise of sufficient accuracy and reliability.

1.3 Holography

1.3.1 Introduction

The second area of interest for this thesis is holography. A traditional optical camera, Figure 1.2, uses an optical chain to focus the scattered light from an object onto a single

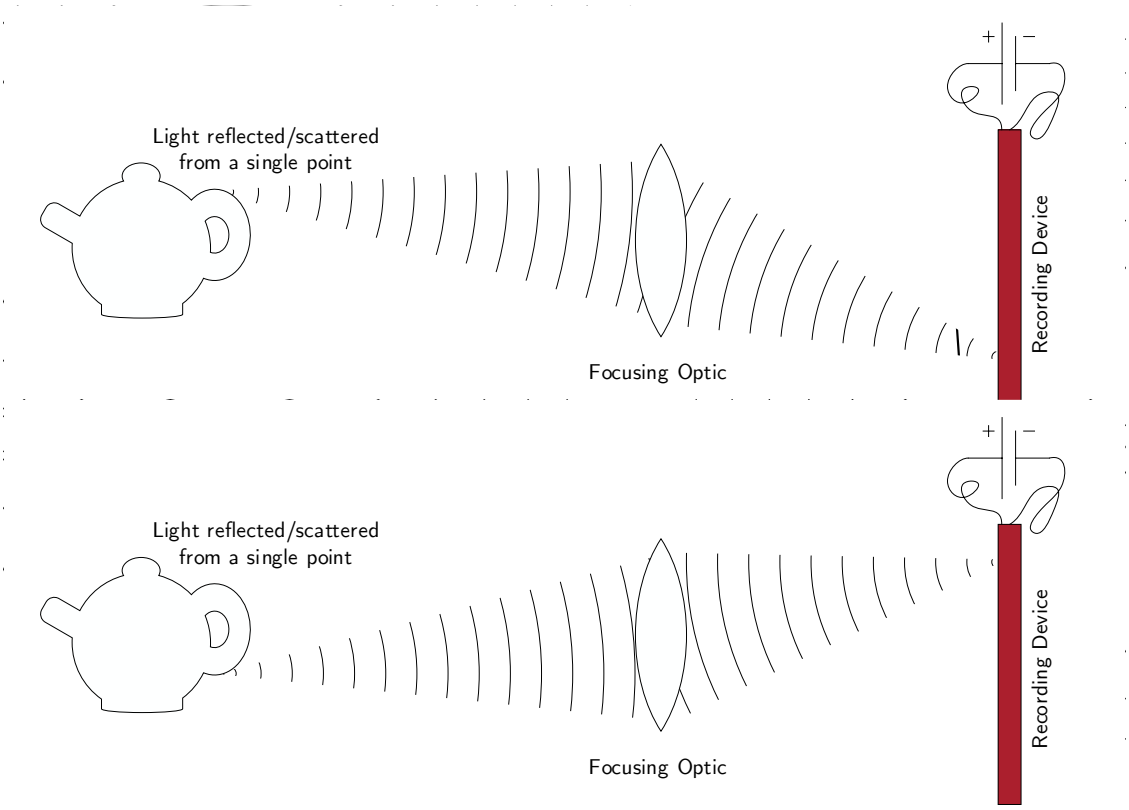


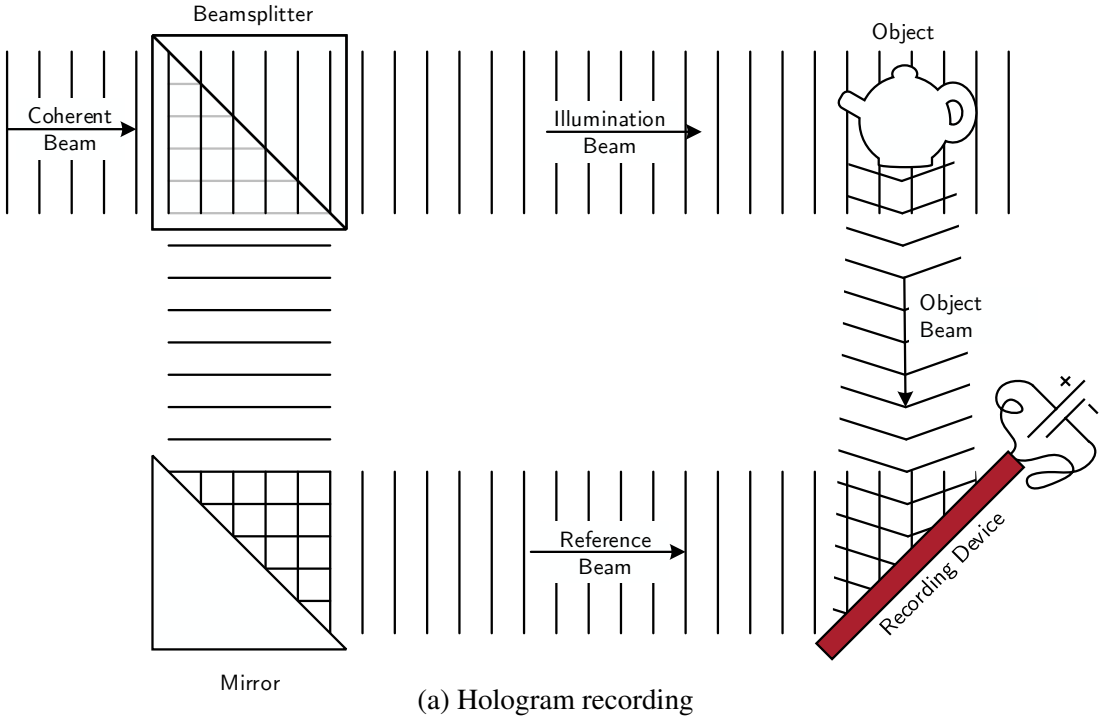
Fig. 1.2 Traditional optical camera

point on a recording device. A difference in location on the recording device corresponds to a difference in approach vector of incident light. Loss of a portion of the record will cause a corresponding loss in the image. A holographic system, on the other hand, collects the scattered light without requiring the use of a focusing optic, Figure 1.3a, instead interfering it with a reference beam. This original image can then be reconstructed by reproducing the recording conditions, Figure 1.3b. The entire image is stored in any one part of the recording device and loss of a portion of the record only causes a loss in resolution[1] of the image. This leads to a regime similar to Figure 1.3.

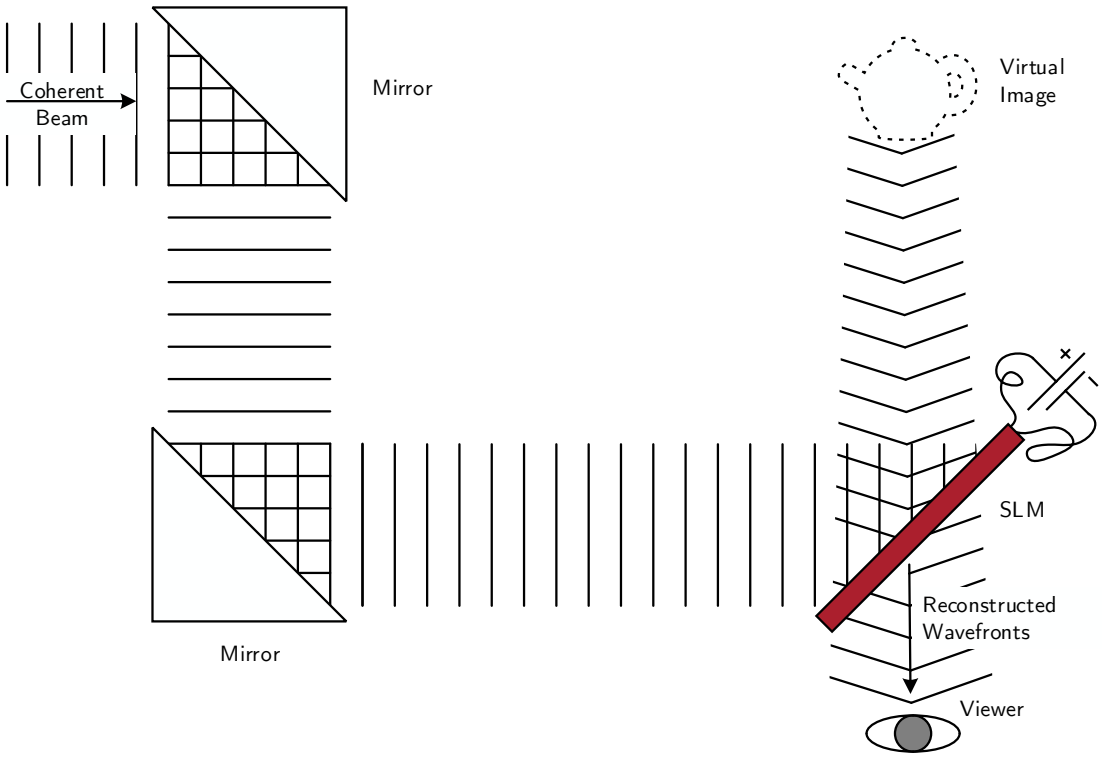
1.3.2 Spatial Light Modulators

Spatial light modulators (SLMs) are integral to holographic systems. The two main areas in SLM research, by volume, are liquid crystals and micromirrors[13]. Many other techniques

[1]Resolution here denotes the ability to *resolve* features in the image rather than referring to sampling density as it is more commonly used when describing a rastered image.



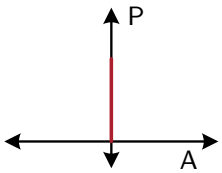
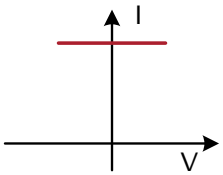
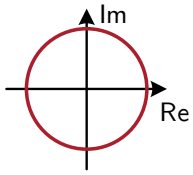
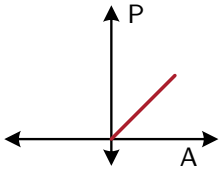
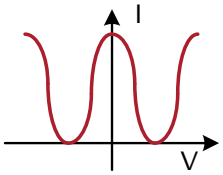
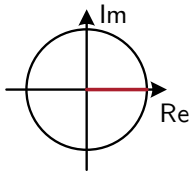
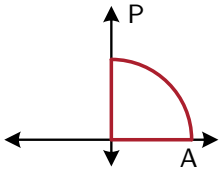
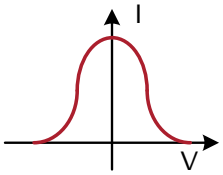
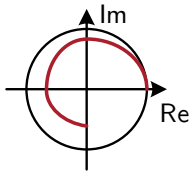
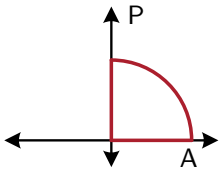
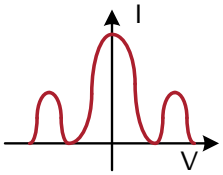
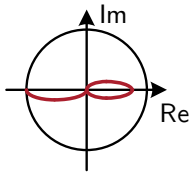
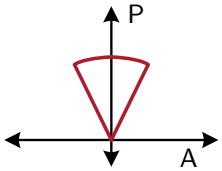
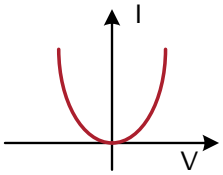
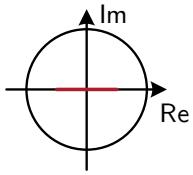
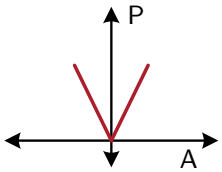
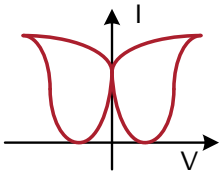
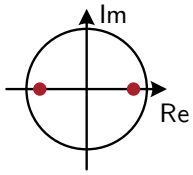
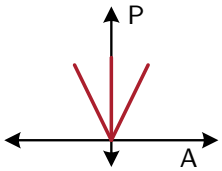
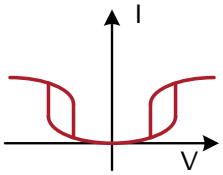
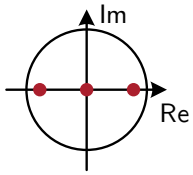
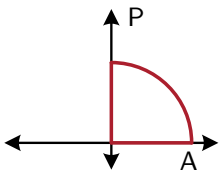
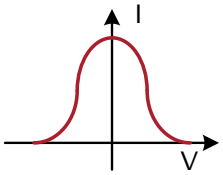
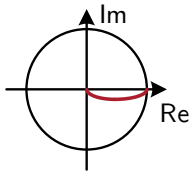
(a) Hologram recording



(b) Hologram reconstruction

Fig. 1.3 Hologram recording and reconstruction

Table 1.1 Types of liquid crystal[12]

LC Type	Optical Configuration	Modulation Range	Complex Modulation	Response Times
Nematic				10-100 µs
Nematic				10-100 µs
Twisted Nematic				20-50 µs
Twisted Nematic				20-50 µs
Smectic A				10-100 µs
Smectic B				10-100 µs
Antiferro-electric				10-100 µs
Twisted SmC				10-100 µs

P - polariser; *A* - analyser; *I* - intensity; *V* - voltage

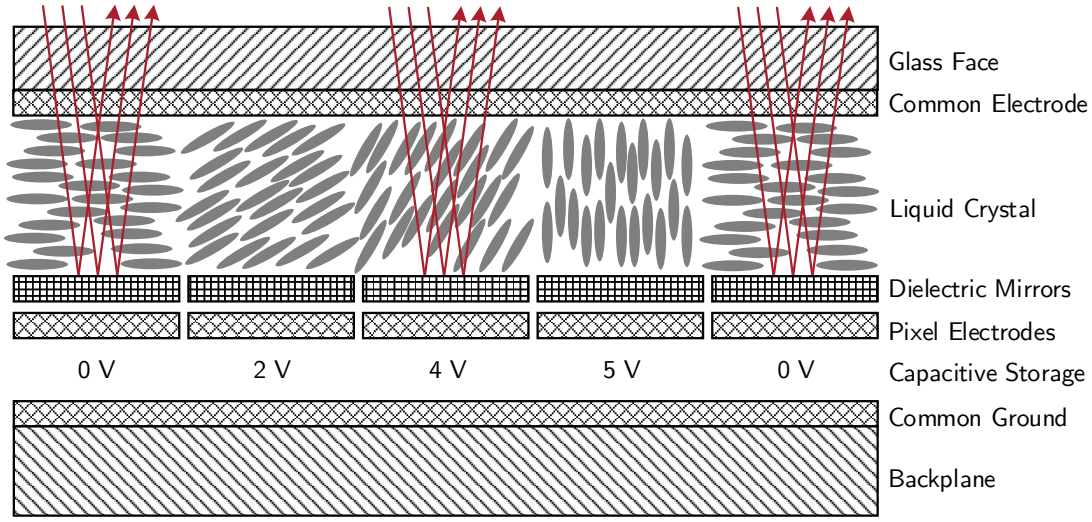


Fig. 1.4 Liquid crystal operation showing the variation in orientation with potential

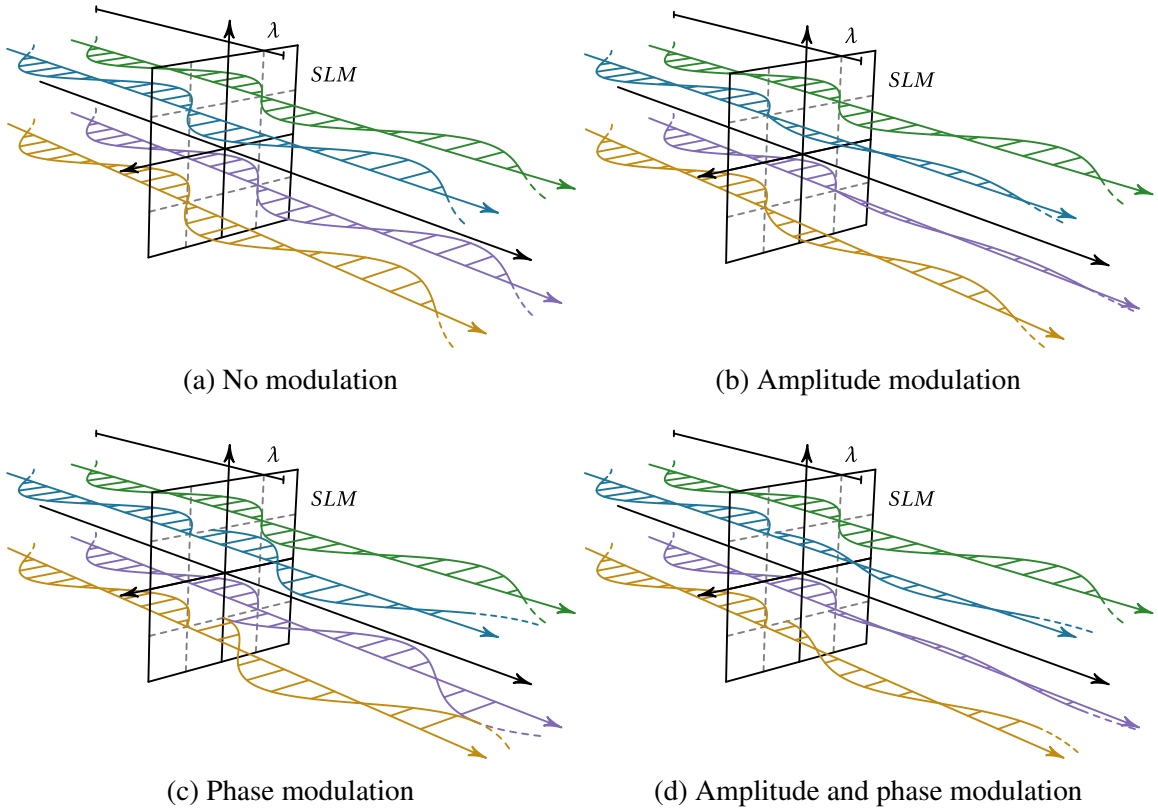
have been researched including viscoelastic modulators; deformable mirrors; crystal based diffraction gratings and microelectromechanical diffraction gratings[14].

Thermotropic liquid crystals (LCs) use an electric field to vary the orientation of the crystals, changing the directional refractive indices, Figure 1.4[15]. This allows for fine control of transmission properties, both spatially and directionally. A major limitation is the requirement for polarised light input and the associated difficulty of finding a sufficiently powerful polarised light source.

Micromirror-based SLMs also emerged in the 1970s and digital micromirror devices (DMDs) developed by Texas Instruments became the leading technology in use[16]. Each individual mirror is bi-stable and able to switch very rapidly between positions. Combining many of these mirrors into an array allows highly detailed raster images to be presented.

DMDs are widely used in laser beam machining and are known to have long lifetimes[17]. By nature, DMDs are prone to significant diffraction effects when operated with a coherent light source.

Depending on wavelength, more than 15% of the incident light is absorbed by the device[18, 19]. For high power applications including projectors, air or liquid cooling is normally incorporated into the DMD housing. The mirrors are typically made out of aluminium though silver, gold and platinum are also used. Most micromirror devices are fabricated using a



(a) No modulation

(b) Amplitude modulation

(c) Phase modulation

(d) Amplitude and phase modulation

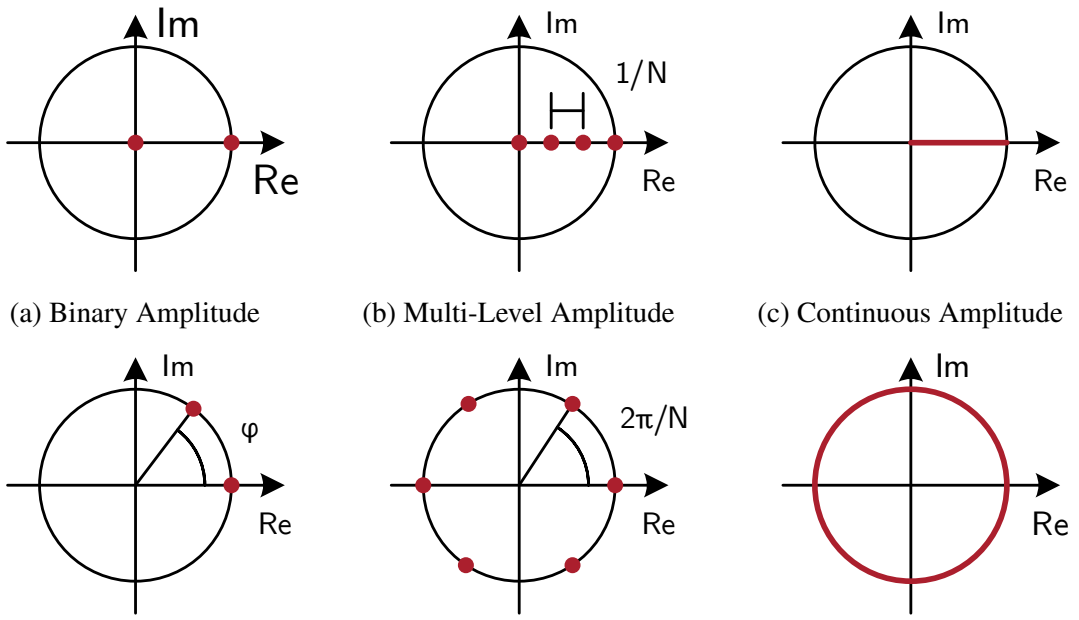
Fig. 1.5 Liquid crystal SLM modulation behaviour

CMOS process with sacrificial layers being used to allow the mirror to be manufactured in place before being removed with a chemical etchant[19].

Another competing technology is the grating light valve (GLV)[20]. GLVs are diffractive micro-optoelectromechanical system (MOEMS). While having a higher bandwidth than comparable DMDs they lack the power capacity and suffer from similar issues.

While DMDs are usually the SLM of choice for high power applications, they are unsuitable in this application due to their exclusively binary with a poor fill factor. While capable of handling a high power load but they also absorb a significant percentage of incident power[21]. By using a phase modulating device with a better fill factor, it will possible to achieve much higher power efficiencies.

There are a wide number of liquid crystal types available and a full discussion is beyond the scope of this work. A summary of the different types of LC is presented in Table 1.1 along with power/voltage response curves and argand diagrams of the modulation schemes[22]. Key quantifiers are whether the SLM modulates the phase or the amplitude of the wave as



(a) Binary Amplitude (b) Multi-Level Amplitude (c) Continuous Amplitude

(d) Binary Phase (e) Multi-Level Phase (f) Continuous Phase

Fig. 1.6 Common modulation schemes

well as whether it is capable of more than binary control. Different modulation behaviours are visualised in Figure 1.5. For example, a nematic LC display is multi-level whereas ferroelectric LCs are merely binary in their operation. The different modulation schemes available are summarised in Figure 1.6.[2]

This project focuses exclusively on LC technologies and for reasons covered in Section 2.2 focuses on multi-level phase modulation. Throughout the rest of the report, SLMs are assumed to be liquid crystal based and the terms are used interchangeably.

1.3.3 Advantages

Holographic projectors have a number of advantages over traditional systems including:

Efficiency - High power projectors typically exhibit less than 10% efficiency with cooling representing the majority of the bulk of modern products. The amplitude modulation element

[2]The terms *modulation* and *quantisation* are often confused and conflated. In this work, modulation is taken to be the input/output curve of the liquid crystal in use and the act of adapting a hologram to meet these constraints. Quantisation is taken similarly with the addition of discrete states or levels found in digital systems. Where the distinction is not relevant the terms are used interchangeably.

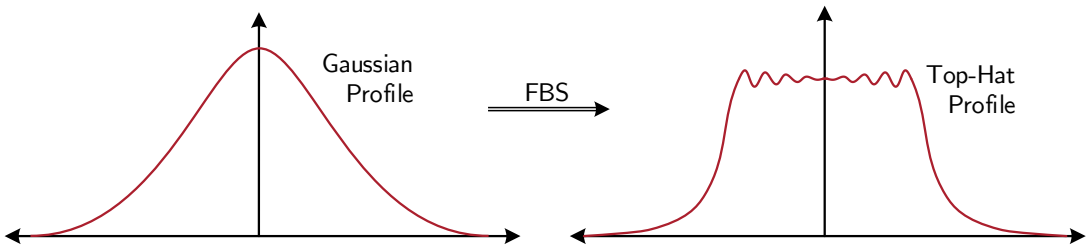


Fig. 1.7 Beam shaping

implies a physical limit to efficiency due to the requirement to block incident light. A typical colour display pixel average at around 20% transmission, the remainder absorbed by the SLM. In contrast, multi-level phase modulators have a very high theoretical efficiency due to absence of conjugate orders and residual zero orders, Section 2.2.2.

Dissipation - The use of phase modulation allows for the redirection of *waste* energy away from the SLM towards heatsinks unlike in traditional systems.

Adaptive Resolution - The continuous nature of holographic images allows for much greater control of the output image. Advanced generation algorithms allow for adaptive local resolutions in areas of interest at the expense of resolution elsewhere.

Robustness, Size and Cost - Holographic systems do not require polarisers, unlike traditional projectors, and far fewer lenses. This helps reduce system size and cost and improve reliability.

1.4 Goals

Traditional AM systems have limited ability to manage or compensate for heat generated. The rapidly moving heat source spot results in high thermal cycling and is a major influence on residual stress and distortion. Mechanical limitations in the galvoscanner mean that over- or under-heating is common and can lead to voids, boiling and spatter. The scale difference between the part size and the spot size means that predictive modelling is beyond the scope of even today's best computing clusters. These factors have led to frequent inability to ensure part quality without physical prototyping and destructive testing.

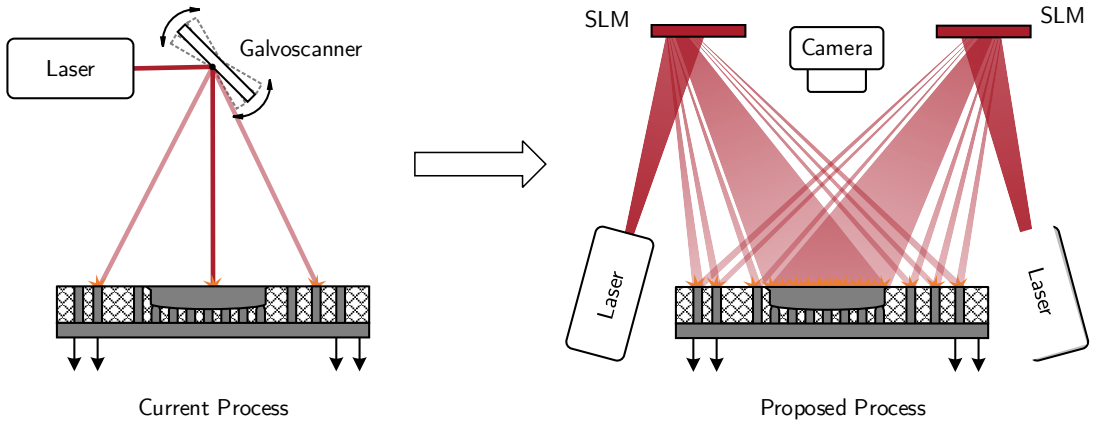


Fig. 1.8 Proposed additive manufacture process compared to traditional processes.

The aim of this PhD has been to create a radically new AM process, holographic additive manufacture (HAM), that uses holographic beam shaping to produce complex light patterns in a single pulse. The differences are summarised in Figure 1.8.

The areal exposure is expected to overcome many of the thermal and material issues currently seen in AM processes. Projecting power to the whole layer at once means that the thermal properties of the powders before and after writing can be factored into the process hologram and part design. It will also significantly reduce thermal gradients and melt-pool instability.

1.5 Previous Syntheses

Somewhat surprisingly, the author is unaware of CGH having ever been proposed for AM, 3DP or RP. That being said, computer-generated holograms have been used in similar applications. Bay pioneered the use of CGH for photolithography demonstrating micrometre features using a 405nm wavelength source [23, 24]. Similarly, Parry [25], used a holographic system to laser mark aluminium and stainless steel.

Perhaps the closest area is that of metal stereolithography where suspensions of metallic powders in resin are printed using resin printing techniques [26, 27]. These approaches suffer from poor final part density and a difficult curing process. This is fundamentally very different from the goals of this thesis as it is only the resin matrix which is cured using an areal approach before the debinding removes the resin matrix and the metallic suspension is sintered.

Arguably the most relevant area of comparible research has been in beam shaping. For example, converting a Gaussian beam to a top hat profile using fundamental beam-mode shaping (FBS) for more even energy input, Figure 1.7[28]. By far the dominant means of doing this is via fixed optical systems using lenses or diffractive elements[29–33]. While there are distinct differences in methodology, complexity and scale, SLMs have been used to produce arbitary beam shapes for high-power lasers[34].

1.6 Summary

This chapter has introduced the fields of additive manufacture and holography. Chapters 2 and Chapter 3 expand on the optical and materials aspects of this respectively. The remaining chapters and appendices then discuss individual aspects of the work carried out in this PhD.

Chapter 2

Mathematical Preliminaries

Chapter 2

Mathematical Preliminaries

This section aims to introduce the reader to the mathematical background required for understanding the material in this thesis. Any required background concepts will be explained where appropriate[1].

2.1 Diffraction

2.1.1 Background

A Fourier transform (FT) decomposes a time based signal in t into a function of the constituent frequencies in ω

$$f(t) \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\rightleftarrows}} F(\omega) \tag{2.1}$$

In one dimension this is represented as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i w t} dt \tag{2.2}$$

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{2\pi i w t} d\omega \tag{2.3}$$

[1]Except where otherwise stated, the mathematical formulae discussed here are taken from[35, 36].

For example

$$f(t) = a\text{Rect}(b,t) \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\rightleftarrows}} F(\omega) = ab\text{sinc}\left(\frac{b\omega}{2}\right) \tag{2.4}$$

where a and b are constant terms, t represents time and ω represents frequency. In two dimensions this becomes

$$F(u,v) = \mathscr{F}\{f_{x,y}\} = \iint_{\pm\infty} f(x,y)e^{-2\pi i(ux+vy)}\,dx\,dy \tag{2.5}$$

$$f(x,y) = \mathscr{F}^{-1}\{F_{u,v}\} = \iint_{\pm\infty} F(u,v)e^{2\pi i(ux+vy)}\,dx\,dy \tag{2.6}$$

where x and y represent the source coordinates and u and v represent the *spatial frequencies* with associated discrete forms.

2.1.2 Principles

The travel of electromagnetic waves obeys Maxwell's equations. In the Heaviside formulation these are[37].

$$\nabla \bullet E = 4\pi\rho \qquad \textit{Coulomb's Law} \tag{2.7}$$

$$\nabla \bullet B = 0 \qquad \textit{Gauss's Law} \tag{2.8}$$

$$\nabla \times E = -\frac{1}{c}\frac{\partial B}{\partial t} \qquad \textit{Faraday's Law} \tag{2.9}$$

$$\nabla \times B = \frac{1}{c}\left(4\pi J + \frac{\partial E}{\partial t}\right) \qquad \textit{Ampère's Law} \tag{2.10}$$

where E is the electric field, B the magnetic field and c the speed of light.

The solution to these equations leads to wave propagation like that shown in Figure 2.3. Figure 2.3a shows the *right-handed* relationship between the electric E and magnetic B fields and Figure 2.3b shows the x and y components of the electric field. Figure 2.3c and Figure 2.3d show the effects of a $\frac{\pi}{2}$ and π delay in the y axis component corresponding to a quarter- and a half- waveplate respectively.

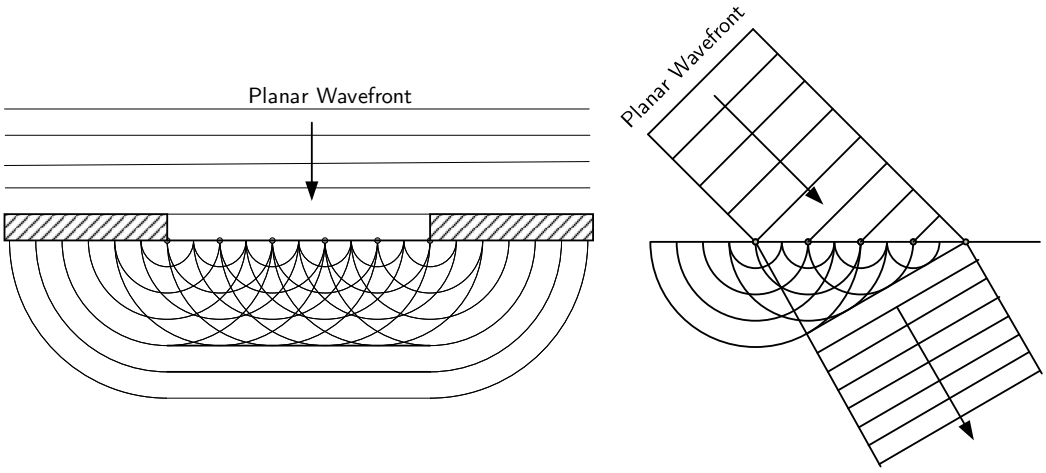


Fig. 2.1 Huygens-Fresnel principle demonstrated during diffraction (left) and refraction (right)

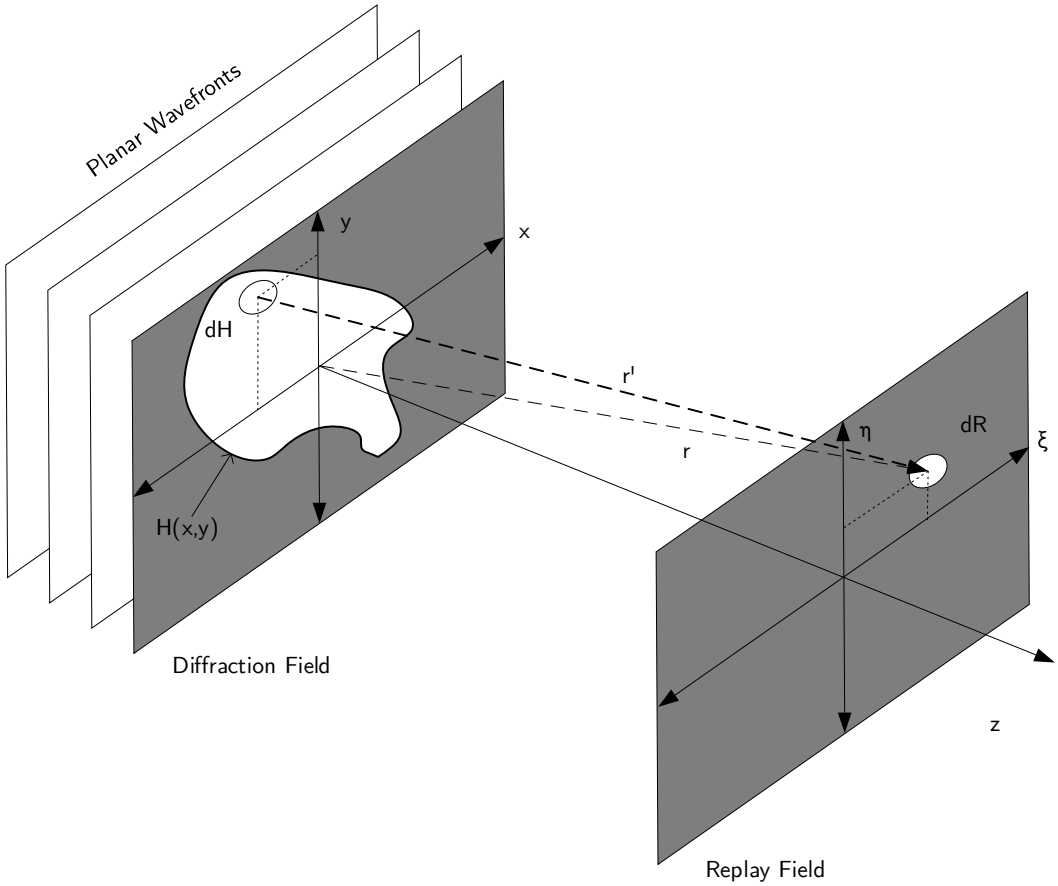
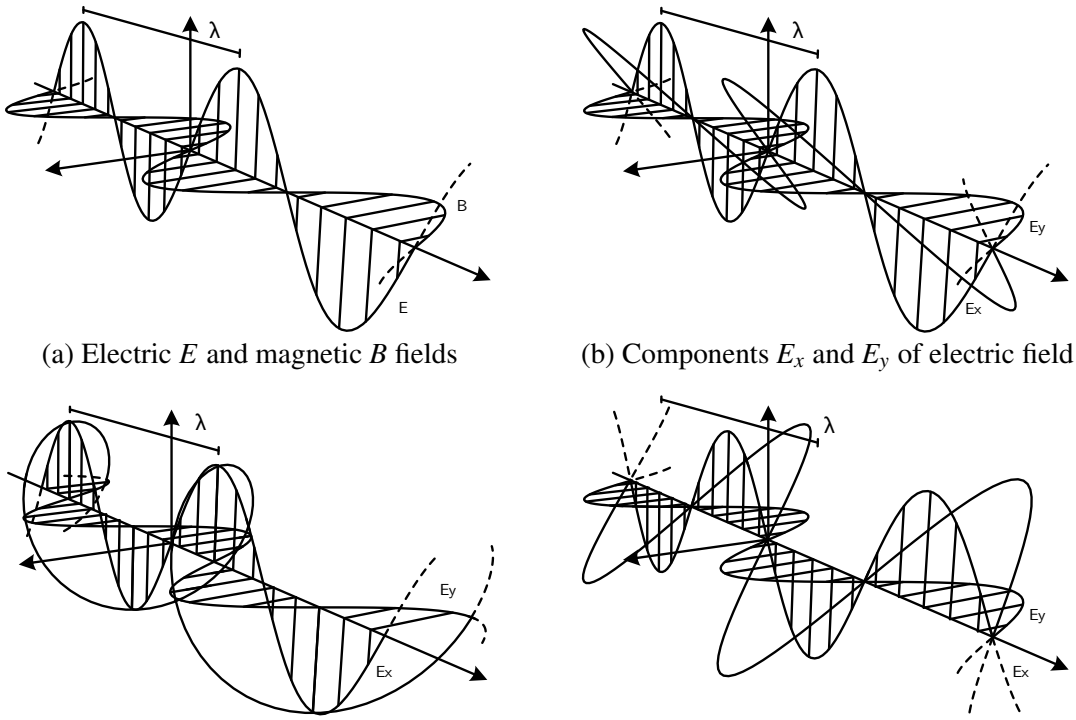


Fig. 2.2 Coordinate systems used in describing a hologram



(a) Electric E and magnetic B fields

(b) Components E_x and E_y of electric field

(c) Effect of $\frac{\pi}{2}$ delay in E_y

(d) Effect of π delay in E_y

Fig. 2.3 Electromagnetic polarisation

Polarisation states can be manipulated using Jones matrices

$$J = \begin{bmatrix} E_x(t) \\ E_y(t) \end{bmatrix} = \begin{bmatrix} E_{x_0} e^{i\phi_x} \\ E_{y_0} e^{i\phi_y} \end{bmatrix} e^{i(\omega t - kz)} \tag{2.11}$$

where E_x and E_y represent the field components in the x and y axes respectively and the $e^{i(\omega t-kz)}$ represents the time propagation. Common Jones matrices for polarisation states and optical components are shown in Table 2.1. The combination of these matrices allows the description of many optical systems.

Pure polarisation states can be represented on a polarisation ellipse (PE) as shown in Figure 2.4. The *erientation angle*, ψ, represents the polarisation angle and the *eccentricity*, ε or ellipticity angle, χ, represent the ratio between proportions.

More complex states can be split into four *Stokes* parameters with S_0 representing the overall intensity, S_1 and S_2 the partial polarisations and S_3 the shape parameter.

Table 2.1 Normalised Jones Matrices for common states and components

Horizontal Polarisation	Vertical Polarisation	±45° Polarisation	Circular RH Polarisation	Circular LH Polarisation
$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\\pm1\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}$

Horizontal Polariser	Vertical Polariser	±45° Polariser	RH Circular Polariser	LH Circular Polariser
$\begin{bmatrix}1&0\\0&0\end{bmatrix}$	$\begin{bmatrix}0&0\\0&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&\pm1\\\pm1&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&i\\-i&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&-i\\i&1\end{bmatrix}$

Horizontal $\frac{\lambda}{4}$ Wave Plate	Vertical $\frac{\lambda}{4}$ Wave Plate	Arbitrary $\frac{\lambda}{4}$ Wave Plate*
$e^{-\frac{i\pi}{4}}\begin{bmatrix}1&0\\0&i\end{bmatrix}$	$e^{\frac{i\pi}{4}}\begin{bmatrix}1&0\\0&-i\end{bmatrix}$	$e^{-\frac{i\pi}{4}}\begin{bmatrix}\cos^2\theta+i\sin^2\theta & (1-i)\sin\theta\cos\theta\\(1-i)\sin\theta\cos\theta & \sin^2\theta+i\cos^2\theta\end{bmatrix}$

Horizontal $\frac{\lambda}{2}$ Wave Plate	Vertical $\frac{\lambda}{2}$ Wave Plate	±45° $\frac{\lambda}{2}$ Wave Plate	Arbitrary $\frac{\lambda}{2}$ Wave Plate*
$e^{-\frac{i\pi}{2}}\begin{bmatrix}1&0\\0&-1\end{bmatrix}$	$e^{\frac{i\pi}{2}}\begin{bmatrix}-1&0\\0&1\end{bmatrix}$	$e^{\frac{i\pi}{2}}\begin{bmatrix}0&\pm1\\\pm1&0\end{bmatrix}$	$e^{\frac{-i\pi}{2}}\begin{bmatrix}\cos2\theta & \sin2\theta\\\sin2\theta & -\cos2\theta\end{bmatrix}$

Arbitrary Birefringent Material*
$e^{-\frac{i\eta}{2}}\begin{bmatrix}\cos^2\theta+e^{i\eta}\sin^2\theta & (1-e^{i\eta})e^{-i\phi}\sin\theta\cos\theta\\(1-e^{i\eta})e^{i\phi}\sin\theta\cos\theta & \sin^2\theta+e^{i\eta}\cos^2\theta\end{bmatrix}$

*Where θ is the fast axis orientation, ϕ is the circularity and η is the relative phase retardation $\phi_y-\phi_x$

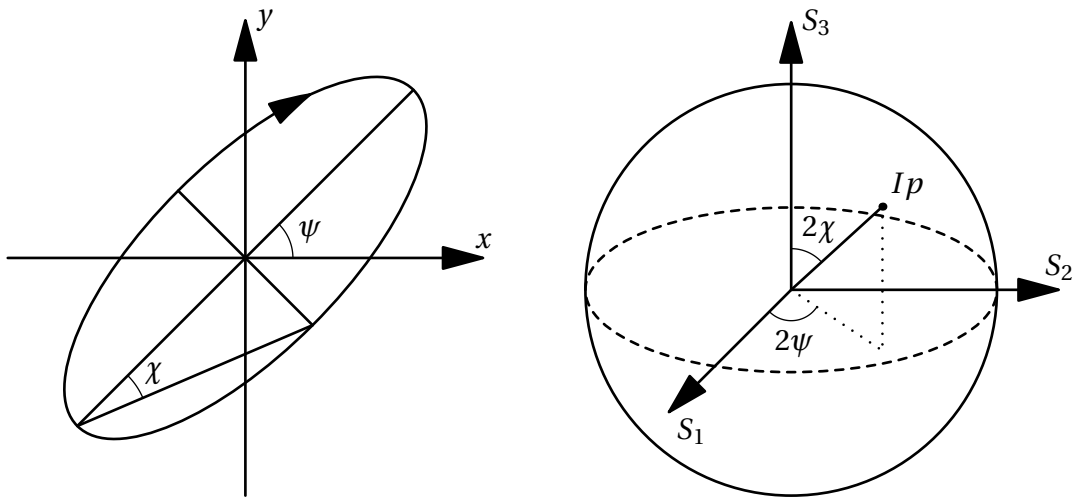


Fig. 2.4 Polarisation ellipse (left) and Poincaré sphere (right)

$$S_0 = I \tag{2.12}$$

$$S_1 = Ip\cos(2\psi)\cos(2\chi) \tag{2.13}$$

$$S_2 = Ip\sin(2\psi)\cos(2\chi) \tag{2.14}$$

$$S_3 = Ip\sin(2\chi) \tag{2.15}$$

where p is the fractional polarisations and I the total intensity. These parameters can be mapped to a *Poincaré* or *Bloch* sphere as shown in Figure 2.4.

In a linear, isotropic and homogeneous medium the Maxwell equations can be reduced to wave propagation equation

$$\nabla^2 u = \frac{n^2}{c^2}\frac{\partial^2 u}{\partial t^2} \tag{2.16}$$

where u is a time-varying vector field and n is the refractive index.

From this, the *Huygens-Fresnel Principle* can be derived which states that every point on the front of an advancing wavelet becomes the source of another spherical wave as seen in Figure 2.1. The angle between the original wave-front and new wavelet θ governs the power of the new wavelet by means of the *obliquity factor*: $1 - cos(\theta)$.

Figure 2.2 shows effect of a point source on the viewing plane or *replay field* $R(u,v)$. For a wave of frequency ω and wave number $k = \frac{2\pi}{\lambda}$, the amplitude or hologram function $H(x,y)$

is given by the following

$$dR = \frac{H(x,y)dS}{r} e^{i(\omega t - kr)} dxdy \tag{2.17}$$

Changing coordinate systems with

$$r' = r\sqrt{1 - \frac{2\xi x + 2\eta y}{r^2} + \frac{x^2 + y^2}{r^2}} \tag{2.18}$$

gives

$$dR = \frac{H(x,y)e^{i\omega t} e^{-ikr\sqrt{1 - \frac{2\xi x + 2\eta y}{r^2} + \frac{x^2+y^2}{r^2}}}}{r\sqrt{1 - \frac{2\xi x + 2\eta y}{r^2} + \frac{x^2+y^2}{r^2}}} dxdy \tag{2.19}$$

Analytical solutions for this are only possible with assumptions valid for different regions of the space around the aperture as in Figure 2.5. When the replay field is sufficiently removed from the source behaviour follows the *far-field* or *Fraunhofer* conditions. Close to the source, *near-field* behaviour is observed. The region in-between is termed the *Fresnel* region.

2.1.3 Fraunhofer Region

In the far-field or *Fraunhofer* region where the aperture is small relative to R and the wavefront can be modelled as planar and the obliquity factor assumed to be negligible, it can be assumed that

$$r \approx r' \tag{2.20}$$

$$r^2 \gg x^2 + y^2 \tag{2.21}$$

Using only the first two terms for the binomial expansion of (2.19) now gives

$$dR = \frac{H(x,y)}{r} e^{i(\omega t - kr)} e^{-\frac{ik}{r}(\xi x + \eta y)} dxdy \tag{2.22}$$

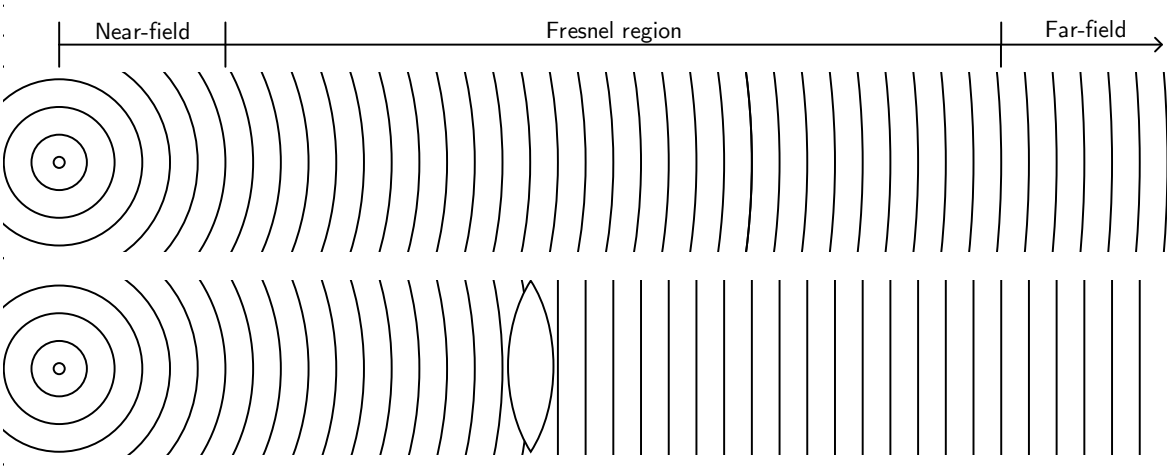


Fig. 2.5 Diffraction regions showing near-field, Fresnel and far-field regions (top) and effect of focusing lens (bottom)

This is equivalent to the *paraxial approximation* which assumes that only rays travelling approximately parallel to the axis need to be considered.

Integrating (2.22) and representing in spatial frequency coordinates gives

$$R(u,v) = \frac{e^{i(\omega t - kr)}}{r} \iint_H H(x,y) e^{-2i\pi(ux+vy)}\, dx\, dy \tag{2.23}$$

This can then be seen as the Fourier transform $R(u,v) \propto \mathscr{F}\{H(x,y)\}$ multiplied by a scale factor where u and v represent the spatial frequency coordinates.

$$u = \frac{\xi k}{2\pi r} \tag{2.24}$$

$$v = \frac{\eta k}{2\pi r} \tag{2.25}$$

which can be discretised for an N_x by N_y grid of pixels

$$R_{u,v} = \frac{e^{i(\omega t - kr)}}{r} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} H_{x,y} e^{-2i\pi(ux+vy)} \tag{2.26}$$

In Cartesian coordinates where the replay field is at distance z from the diffraction field this becomes

$$R(\xi,\eta) = \underbrace{\frac{e^{\frac{i\pi}{\lambda z}(2z^2+\xi^2+\eta^2)}}{i\lambda z}}_{\text{Scale Factor}} \underbrace{\iint_H H(x,y)e^{-\frac{2i\pi}{\lambda z}(\xi x+\eta y)}\,dx\,dy}_{\mathscr{F}\{H(x,y)\}} \tag{2.27}$$

which can also be discretised for an N_x by N_y grid of pixels

$$R(\xi,\eta) = \underbrace{\frac{e^{\frac{i\pi}{\lambda z}(2z^2+\xi^2+\eta^2)}}{i\lambda z}}_{\text{Scale Factor}} \underbrace{\sum_{x=0}^{N_x-1}\sum_{y=0}^{N_y-1} H(x,y)e^{-\frac{2i\pi}{\lambda z}(\frac{\xi x}{N_x}+\frac{\eta y}{N_y})}}_{DFT\{H_{x,y}\}} \tag{2.28}$$

Goodman suggests the following criteria for the start of the Fraunhofer region

$$r \gg \frac{\pi(\Delta_x^2+\Delta_y^2)}{\lambda} \tag{2.29}$$

where Δ_x and Δ_y are the x and y *pixel pitches* respectively. The *antenna designer's formula* gives a lower estimate where

$$r > \frac{2\sqrt{\Delta_x^2+\Delta_y^2}}{\lambda} \tag{2.30}$$

In many applications, focusing optics can be used to shorten the value of r required to operate in the far-field - Figure 2.5.

2.1.4 Fresnel Region

The assumption of wavefront planarity no longer holds in the *Fresnel region* and instead a parabolic shape is assumed.

$$r \approx \sqrt{z^2+(x-\xi)^2+(y-\eta)^2} \tag{2.31}$$

Expanding this with the first two terms of a binomial expansion gives

$$dR = \frac{H(x,y)}{r} e^{i(\omega t - kr)} e^{\frac{ik}{2z}(x^2+y^2)} e^{-\frac{ik}{r}(\xi x+\eta y)} dxdy \tag{2.32}$$

$$R(u,v) = \frac{e^{i(\omega t-kr)}}{r} \iint_H H(x,y) e^{\frac{ik}{2z}(x^2+y^2)} e^{-2i\pi(ux+vy)}\, dx\, dy \tag{2.33}$$

In Cartesian coordinates

$$R(\xi,\eta) = \underbrace{\frac{1}{i\lambda z} e^{\frac{2i\pi z}{\lambda}} e^{\frac{i\pi}{\lambda z}(\xi^2+\eta^2)}}_{\text{Scale Factor}} \underbrace{\iint_H H(x,y) \overbrace{e^{\frac{i\pi}{\lambda z}(x^2+y^2)}}^{\text{Quadratic Phase Factor}} e^{-\frac{2i\pi}{\lambda z}(\xi x+\eta y)}\, dx\, dy}_{\mathscr{F}\{H(x,y)e^{\frac{i\pi}{\lambda z}(x^2+y^2)}\}} \tag{2.34}$$

This can be discretised for an N_x by N_y grid of pixels

$$R_{\xi,\eta} = \underbrace{\frac{1}{i\lambda z} e^{\frac{2i\pi z}{\lambda}} e^{\frac{i\pi}{\lambda z}(\xi^2+\eta^2)}}_{\text{Scale Factor}} \underbrace{\sum_{x=0}^{N_x-1}\sum_{y=0}^{N_y-1} H_{x,y} \overbrace{e^{\frac{i\pi}{\lambda z}(x^2+y^2)}}^{\text{Quadratic Phase Factor}} e^{-\frac{2i\pi}{\lambda z}(\frac{\xi x}{N_x}+\frac{\eta y}{N_y})}}_{DFT\{H_{x,y}e^{\frac{i\pi}{\lambda z}(x^2+y^2)}\}} \tag{2.35}$$

Fresnel holograms can be seen to be similar to Fraunhofer holograms with the addition of the *quadratic phase term*

$$e^{\frac{i\pi}{\lambda z}(x^2+y^2)} \tag{2.36}$$

Different formulations can be found for computing Fresnel holograms, though the performance is worse than the Fraunhofer case and requires greater system knowledge.

This project is not concerned with behaviour in the near-field region. This can be thought of as similar to the Fraunhofer case where the paraxial approximation incorporates a second term in the Taylor series expansions of trigonometric identities.

2.1.5 Properties

While the properties of the Fourier transform are well known, a couple deserve mention here.

Parseval's Theorem states that the total energy in the spatial frequency domain is equal to the total energy in the object domain,

$$\int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty}|H(x,y)|^2dxdy=\int_{u=-\infty}^{\infty}\int_{v=-\infty}^{\infty}|R(u,v)|^2dudv \tag{2.37}$$

This can be thought of as a special case of *Plancherel's Theorem*

$$\int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty}G(x,y)\overline{H(x,y)}dxdy=\int_{u=-\infty}^{\infty}\int_{v=-\infty}^{\infty}T(u,v)\overline{R(u,v)}dudv \tag{2.38}$$

where

$$G \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\rightleftarrows}} T, \quad H \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\rightleftarrows}} R \tag{2.39}$$

The second thing worthy of notice is the shift theorem. This states that a translation of u_0 and v_0 in the replay corresponds to a phase shift in the

$$\mathscr{F}\left\{H(x,y)e^{-2\pi i(xu_0+yv_0)}\right\}=R(u-u_0,v-v_0) \tag{2.40}$$

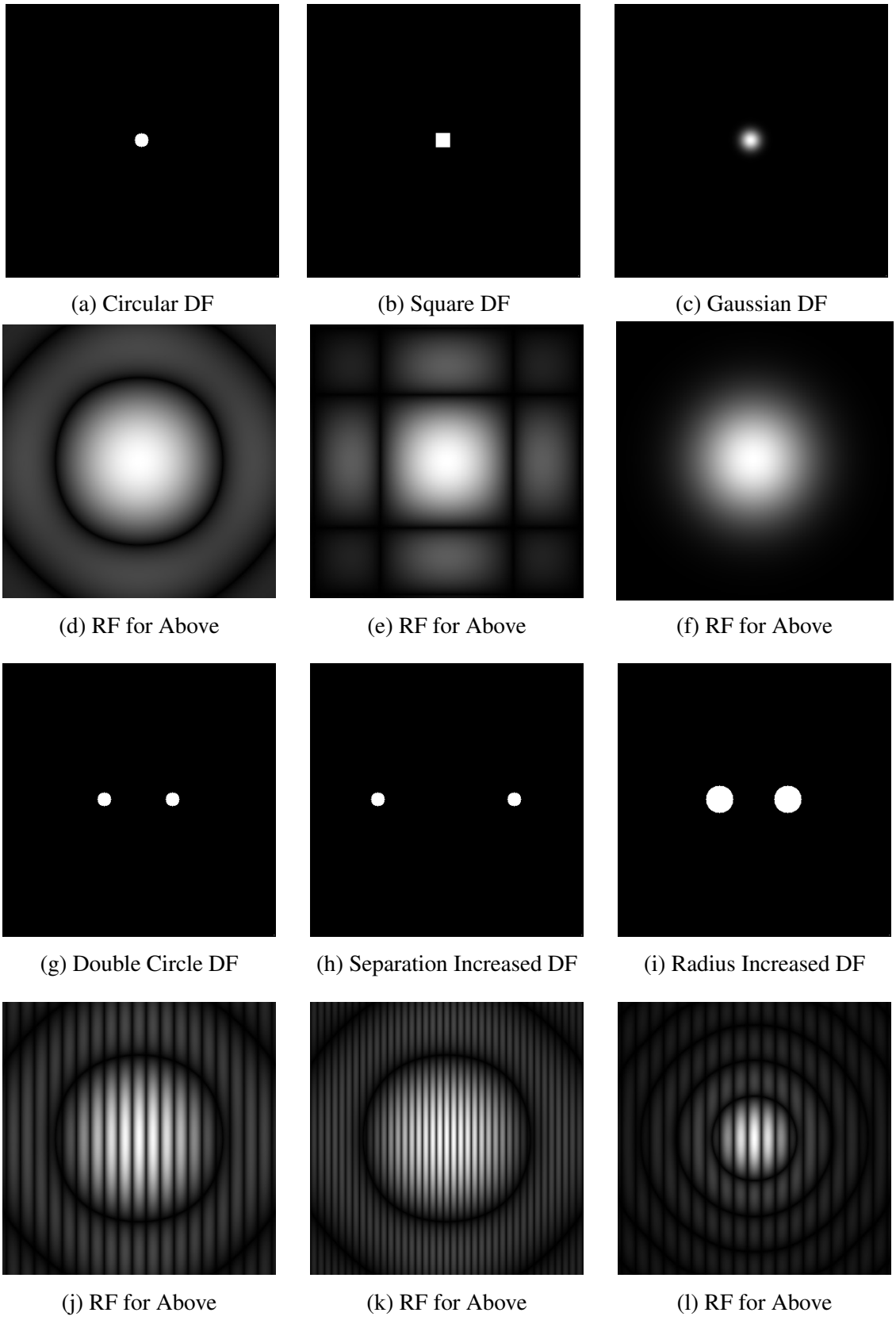
In practical terms, this means that a phase change in one plane is equal to a change of position in the other.

Finally, a result that will be of use later, the convolution theorem

$$\mathscr{F}\{H_1(x,y)H_2(x,y)\}=R_2(u,v)*R_2(u,v) \tag{2.41}$$

$$\mathscr{F}\{H_1(x,y)*H_2(x,y)\}=R_2(u,v)R_2(u,v) \tag{2.42}$$

This allows for the easy combination of existing holograms.



(a) Circular DF (b) Square DF (c) Gaussian DF

(d) RF for Above (e) RF for Above (f) RF for Above

(g) Double Circle DF (h) Separation Increased DF (i) Radius Increased DF

(j) RF for Above (k) RF for Above (l) RF for Above

Fig. 2.6 replay field (RF) patterns generated by given diffraction fields (DF)

2.1.6 Examples

Figures 2.6a, 2.6b and 2.6c show the following gratings lit by a beam of intensity I,

$$H_{square}(x,y) = \begin{cases} I_0(x,y), & |x| < R \text{ and } |y| < R, \\ 0, & \text{otherwise.} \end{cases} \tag{2.43}$$

$$H_{circular}(x,y) = \begin{cases} I_0(x,y), & x^2+y^2 < R^2, \\ 0, & \text{otherwise.} \end{cases} \tag{2.44}$$

$$H_{gaussian}(x,y) = I_0(x,y)e^{-\frac{x^2+y^2}{2r^2}} \tag{2.45}$$

The corresponding replay field patterns, Figures 2.6d, 2.6e and 2.6f, are

$$H_{square}(x,y) = I_0(u,v)sinc(\pi ur)sinc(\pi vr) \tag{2.46}$$

$$H_{circular}(x,y) = I_0(u,v)\frac{J_1(2\pi\sqrt{(u^2+v^2)})}{\sqrt{(u^2+v^2)}} \tag{2.47}$$

$$H_{gaussian}(x,y) = r\sqrt{2\pi}I_0(u,v)e^{8r^2(u^2+v^2)} \tag{2.48}$$

where J_i is the i'th Bessel function and defines the form of an *Airy disc*.

$$J_i(a) = \sum_{n=1}^{\infty} \frac{(-1)^n}{m!\Gamma(n+i+1)} \left(\frac{a}{2}\right)^{2n+i} \tag{2.49}$$

The shift theorem, Section 2.1.5, for two discs Figures 2.6g, 2.6h and 2.6i show three diffraction fields (DFs). Figure 2.6j shows the first case replay field while Figure 2.6k shows the effect of moving the apertures apart, an increase in spatial frequency. An increase in spot radius in Figure 2.6l causes a scale effect on the resulting image.

2.2 Practical Considerations

2.2.1 Image Conjugate

With the exception of multi-level phase modulating SLMs - Section 1.3.2 - all systems produce a replay field with a real valued or DC component. The conjugate symmetry of the FTs underlying both Fraunhofer (2.23) and Fresnel holograms (2.32) mean that unless multiple phase control is available, images are produced with a 180° rotational symmetry as shown in Figure 2.7.

$$R(u,v) = \mathscr{F}\{H(x,y)\} = \mathscr{F}\{\overline{H(-x,-y)}\} = R(-u,-v) \tag{2.50}$$

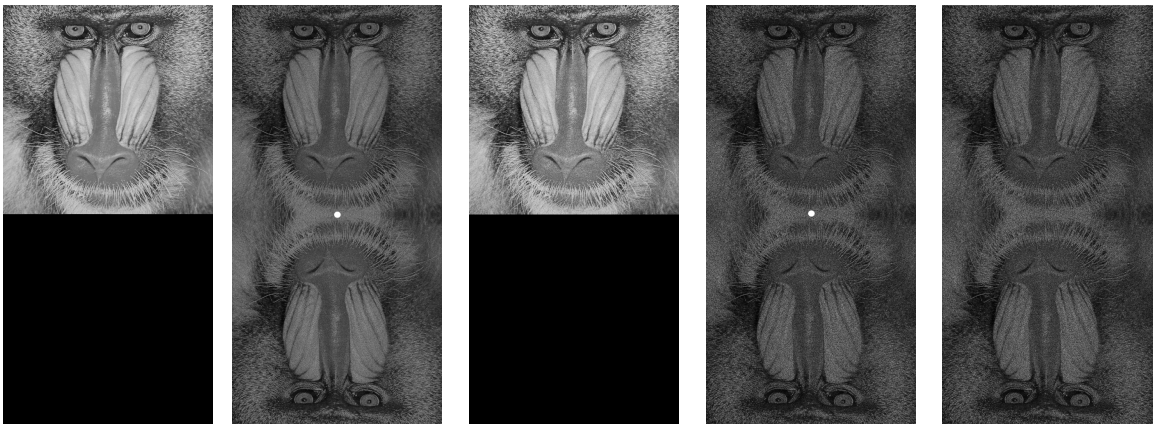
2.2.2 Zero-Order

A zero-order region at the origin is common and is due to undiffracted light. As the replay field is the frequency spectrum of the hologram, the zero-order represents the average pixel value of the image. For amplitude only images, this means that the zero order presence is unavoidable. It can be shown that the zero-order takes 50% of the image power. Phase modulation allows for *DC-balancing* and the removal of the zero order. The effects of zero order and the image conjugate can be seen in Figure 2.7.

2.2.3 Edge-Enhancement

A common issue is that of edge enhancement. The use of a uniform pixel size equalises the power spectrum of a displayed image due to the regular sampling interval. Energy is concentrated at lower wavelengths in many images leading to an enhancement of higher spatial frequencies, amplifying any sharp edges. This high-pass filtering can be compensated for by adding a random phase to each pixel, a process referred to as phase randomisation (PR).

$$H'(x,y) = H(x,y)e^{i\times Rand[0,2\pi]} \tag{2.51}$$



(a) Target (b) Multi-level Amplitude (c) Multi-level Phase (d) Binary Amplitude (e) Binary Phase

Fig. 2.7 Conjugate image and zero-order as seen in different modulation schemes.

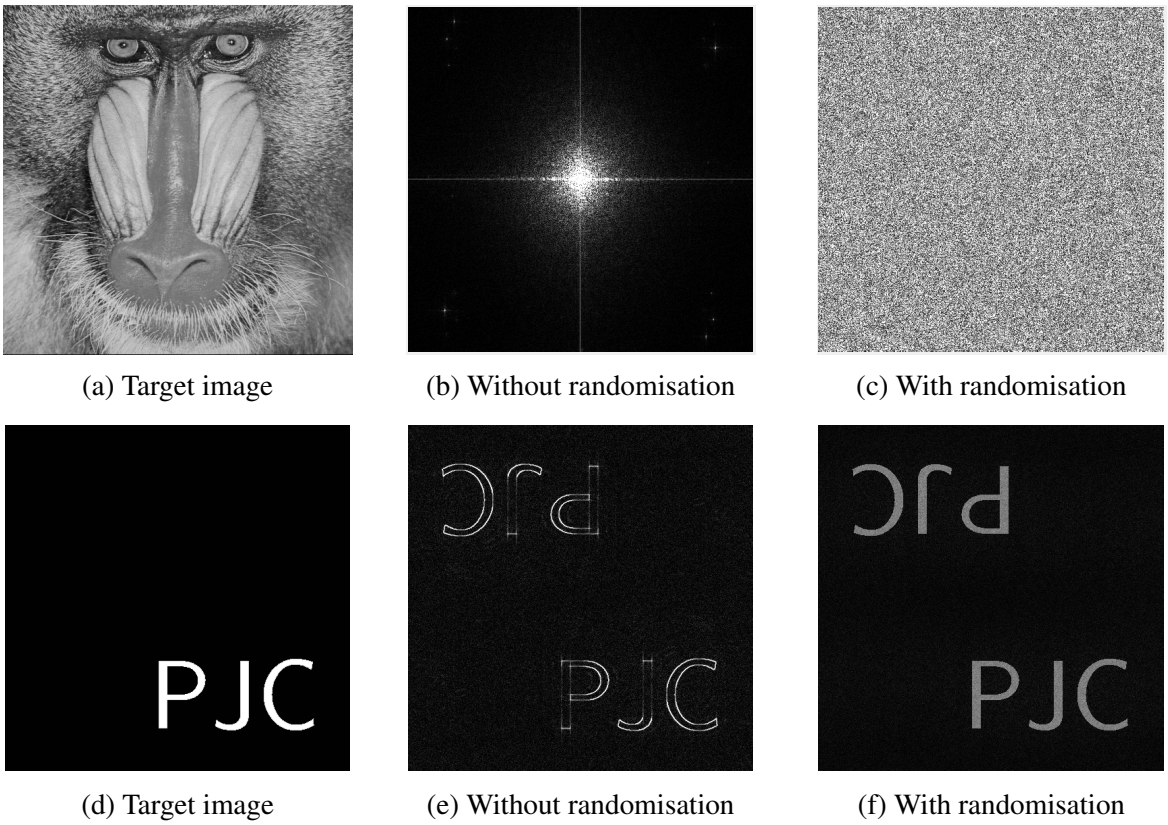
The human eye is phase insensitive so this does not effect image quality and has the effect of smoothing the power spectral density (PSD) of the target. This is demonstrated in Figure 2.8. Similar techniques have also been applied in three-dimensional holography[38].

2.2.4 Coherence and Noise

Another potential issue is caused by the coherence length of the source. For a long coherence length, any noise is clearly reproduced and amplified. Increasing the bandwidth has the effect of smoothing the noise in the image and introducing blurring.

2.2.5 Speckle

Another common issue is that of *speckle*. When viewing a collimated light source against a diffuse background or rough surface, highly volatile and unpredictable interference patterns are produced by the scattered light. Speckle is distinct from other noise and error sources as it is produced at the plane of the observer. This makes the speckle pattern dependent on the viewer and impossible to compensate for. As in the Section 2.2.4, decreasing the coherence length of the source reduces the speckle at the expense of additional image blurring [39]. An alternative approach is to use time-multiplexing where the speckle of multiple frames is averaged.



(a) Target image (b) Without randomisation (c) With randomisation

(d) Target image (e) Without randomisation (f) With randomisation

Fig. 2.8 Effect of phase randomisation on frequency magnitude spectrum (top) and sharp edges (bottom) for a binary phase hologram

2.2.6 Viewing Angle

SLMs exhibits viewing angles θ_x and θ_y which are given as functions of the pixel pitch Δ_x and Δ_y and wavelength λ

$$\theta_x = 2\sin^{-1}\left(\frac{\lambda}{2\Delta_x}\right), \quad \theta_y = 2\sin^{-1}\left(\frac{\lambda}{2\Delta_y}\right) \tag{2.52}$$

With a small angle approximation, a relationship for limiting resolution $\theta_{x,\text{res}}$ and $\theta_{y,\text{res}}$ is obtained from resolution N_x and N_y

$$\theta_{x,\text{res}} = \frac{\lambda}{N_x\Delta}, \quad \theta_{y,\text{res}} = \frac{\lambda}{N_y\Delta} \tag{2.53}$$

2.2.7 Replay Field Size

For a hologram created using a lens of focal length f, the viewing angle is relationship is equivalent to the following relationship for replay size

$$\xi_{\max} - \xi_{\min} = \frac{f\lambda}{2\Delta_x}, \quad \eta_{\max} - \eta_{\min} = \frac{f\lambda}{2\Delta_y} \tag{2.54}$$

where Δ_x and Δ_y are the x and y pixel pitches and $\xi_{\max} - \xi_{\min}$ and $\eta_{\max} - \eta_{\min}$ are the ξ and η replay field dimensions.

2.2.8 Spectral Bandwidth

Assuming a small angle identity, Section 2.2.6 states that the replay field size is proportional to wavelength. The *Spectral Bandwidth* (SB) of a light source is taken as being the Full-Width Half-Maximum (FWHM) bandwidth of the wavelength power spectral density (PSD). For a laser diode this is typically 2-5 nm while for a LED, 20-60 nm is common.

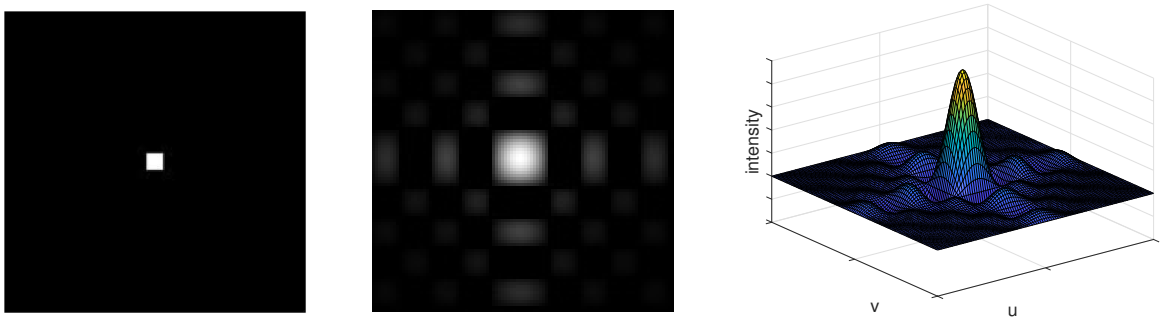
Following Freeman, (2.52) can be extended to give a formulation for the*chromatic spread*, the variation in pixel replay field position due to the spread in diffraction angles caused by the variation in the light wavelength[40, 41]. This effect is most prominent at the outside of replay field. Assuming a minimum SLM feature size of twice the pixel pitch and with a SB of $\delta\lambda$ the spread equals

$$\theta_{x,\text{chrom}} = \frac{\Delta_\lambda}{2\Delta_x}, \quad \theta_{y,\text{chrom}} = \frac{\Delta_\lambda}{2\Delta_y} \tag{2.55}$$

Comparing with (2.53) gives a limit for spectral bandwidth $\Delta_{\lambda,\text{lim}}$ before resolution is effected.

$$\Delta_{\lambda,\text{lim}} = \frac{2\lambda}{\max(N_x, N_y)} \tag{2.56}$$

For the Jasper JD8714 used for much of this thesis, Table 6.1, this gives $\Delta_{\lambda,\text{lim}} = 0.2$nm at $\lambda = 405$.



(a) SLM Pixel (b) Replay Field Map (c) Replay Field Intensity

Fig. 2.9 Rectangular pixel (left) with replay field (centre) and intensity map (right)

2.2.9 Source Size

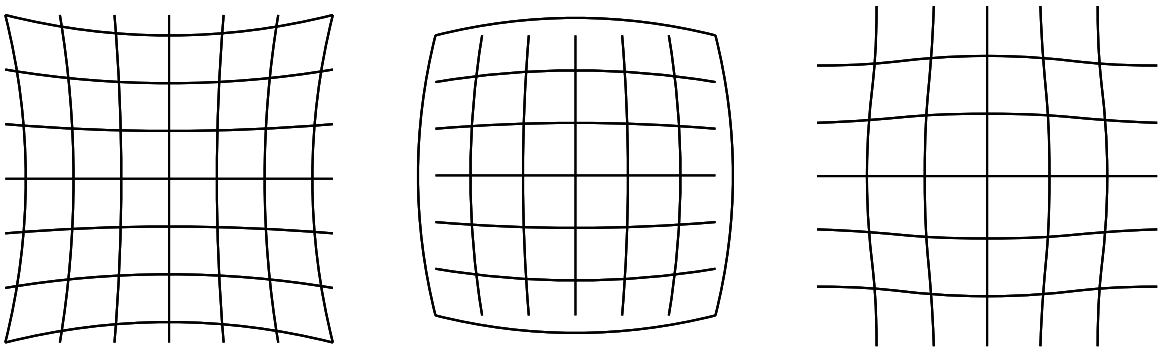
Similarly, the size of the source ν_{src} also effects the spread θ_{src} of the replay field. This can be compared to the limiting resolution (2.53) to give an upper bound on the size of the source.

2.2.10 Super Resolution

The replay field of a computer-generated hologram is a continuous function even if generated by a pixellated diffraction field. The eye only being sensitive to the square of the intensity, $\left|R(x,y)^2\right|$, a hologram without conjugate image has the same number of sample points in the RF as can be modulated in the DF. Nyquist sampling theory suggests that increasing the number of sampling points in the RF will not improve the quality of the RF image[42]. The RF sampling pattern can be modified, however, to improve resolution in portions of the image at the expense of other areas. In the case of binary holograms with conjugate images, Cable makes a case that twice the RF sampling points can be taken before the Nyquist limit is reached[43].

2.2.11 Variable Resolution and Supersampling

While the FFT is conducted on a regular sampling grid, the assessment metrics are not so limited. As a result, it is possible to supersample a portion of the image without sacrificing overall quality provided the number of sampling points remains constant. This follows a similar argument to that in Section 2.2.10.



(a) Pincushion (b) Barrel (c) Mustache

Fig. 2.10 Common distortion types

2.2.12 Pixel Shape Compensation

Due to the need for control electronics and insulating boundaries, the SLM pixels only fill a certain percentage of the display. Assuming that the pixels are square in shape, a *sinc* shaped replay field is expected, Figure 2.9. For an N_x by N_y image, the relative attenuation can be given by

$$\Upsilon(u,v) = \left(\frac{sin\left(\frac{\pi(x-N_x/2)}{M}\right)}{\pi(x-N_x/2)} \cdot \frac{sin\left(\frac{\pi(y-N_y/2)}{M}\right)}{\pi(y-N_y/2)} \right) \tag{2.57}$$

To compensate for this, the amplitude of the target image pixel can be reduced by a factor of $R(u,v) = 1/\sqrt{(\Upsilon(u,v))}$. This results in a lowering of the available intensity meaning that it is preferable to not use the edges of the replay field.

2.3 Distortion and Aberration

Geometric deformations of the image are referred to as *distortions* and effect only the shape of the image[44]. Provided the optical system is cylindrically symmetric, these distortions are easy to characterise and correct for. Common distortion types include the barrel and pincushion distortions shown in Figure 2.10 and can be handled by classical distortion theory or as a subset of the subject of Zernike polynomials[45].

Table 2.2 Zernike polynomial definitions[44]

	Index			Polynomial	
Aberration Description	j	n	m	Radial	Angular
Piston	0	0	0	1	
Tip	1	1	1	ρ	$\cos(\theta)$
Tilt	2	1	-1	ρ	$\sin(\theta)$
Defocus	3	1	0	$2\rho^2-1$	
Astigmatism at 0°	4	2	2	ρ^2	$\cos(2\theta)$
Astigmatism at 90°	5	2	-2	ρ^2	$\sin(2\theta)$
Coma at 0°	6	3	1	$\rho(3\rho^2-2)$	$\cos(\theta)$
Coma at 45°	7	3	-1	$\rho(3\rho^2-2)$	$\sin(\theta)$
Spherical	8	2	0	$6\rho^4-6\rho^2+1$	
Trefoil at 0°	9	3	3	$\rho 3$	$\cos(3\theta)$
Trefoil at 45°	10	3	-3	$\rho 3$	$\sin(3\theta)$
Secondary Astigmatism at 0°	11	3	2	$\rho^2(4\rho^2-3)$	$\cos(2\theta)$
Secondary Astigmatism at 90°	12	3	-2	$\rho^2(4\rho^2-3)$	$\sin(2\theta)$
Secondary Coma at 0°	13	3	1	$\rho(10\rho^4-12\rho^2+3)$	$\cos(\theta)$
Secondary Coma at 45°	14	3	-1	$\rho(10\rho^4-12\rho^2+3)$	$\sin(\theta)$
Secondary Spherical	15	3	0	$20\rho^6-30\rho^4+12\rho^2-1$	

Changes in the phase of the wave caused by imperfect optical components are referred to as *aberrations*. Aberrations can be characterised by the use of Zernike polynomials.

2.3.1 Zernike Polynomials

Zernike polynomials form a complete orthonormal set on the unit circle and are defined in terms of radial and angular coordinates. Each polynomial is given by

$$Z_n^m(\rho,\theta) = N_n^m \times V_n^m(\rho) \times G_n^m(\theta) \tag{2.58}$$

where n is the polynomial order, m the azimuthal frequency, N_n^m the normalisation constant, $V_n^m(\rho)$ the radial polynomial and $G^m(\theta)$ the angular polynomial.

The coefficients are taken as:

$$V_n^m(\rho) = \sum_{s=0}^{\frac{n-m}{2}} \frac{(-1)^s (n-s)!}{s!(\frac{n+m}{2})!(\frac{n-m}{2}-s)!} \rho^{n-2s} \tag{2.59}$$

$$G^m(\theta) = \begin{cases} cos(m\theta), & m >= 0, \\ sin(m\theta), & m < 0. \end{cases} \tag{2.60}$$

$$N_n^m = \sqrt{\frac{2(n+1)}{1+\delta_0(m)}} \tag{2.61}$$

Wyant introduced a single index counting mechanism - $Z_n^m \rightarrow Z_j$ - for Zernike polynomials[46]. Any given correction $C(x,y)$ can be given to arbitrary precision by n polynomials with weights w_j.

$$C_{x,y} = \sum_{j=1}^{n} w_j Z_{z,x,y} \tag{2.62}$$

Table 2.2 shows the first 15 Zernike polynomials while Figure 2.11 provides a graphical reference.

2.3.2 Aberration Correction

Optical system aberrations depend generally on both the image and aperture coordinates and are referred to as field-dependant or spatially varying aberrations[47]. Field-independent aberrations are a special case and are significantly easier to handle.

2.3.3 Field-Independent Aberration Correction

When operating in an imperfect system, the Fraunhofer form of the hologram, $E_{u,v} = \mathscr{F}\{A_{x,y}\}$, can be taken as

$$E_{u,v} = \mathscr{F}\{I_{x,y} H_{\text{uncor},x,y} e^{2\pi i \Phi_{x,y}}\} \tag{2.63}$$

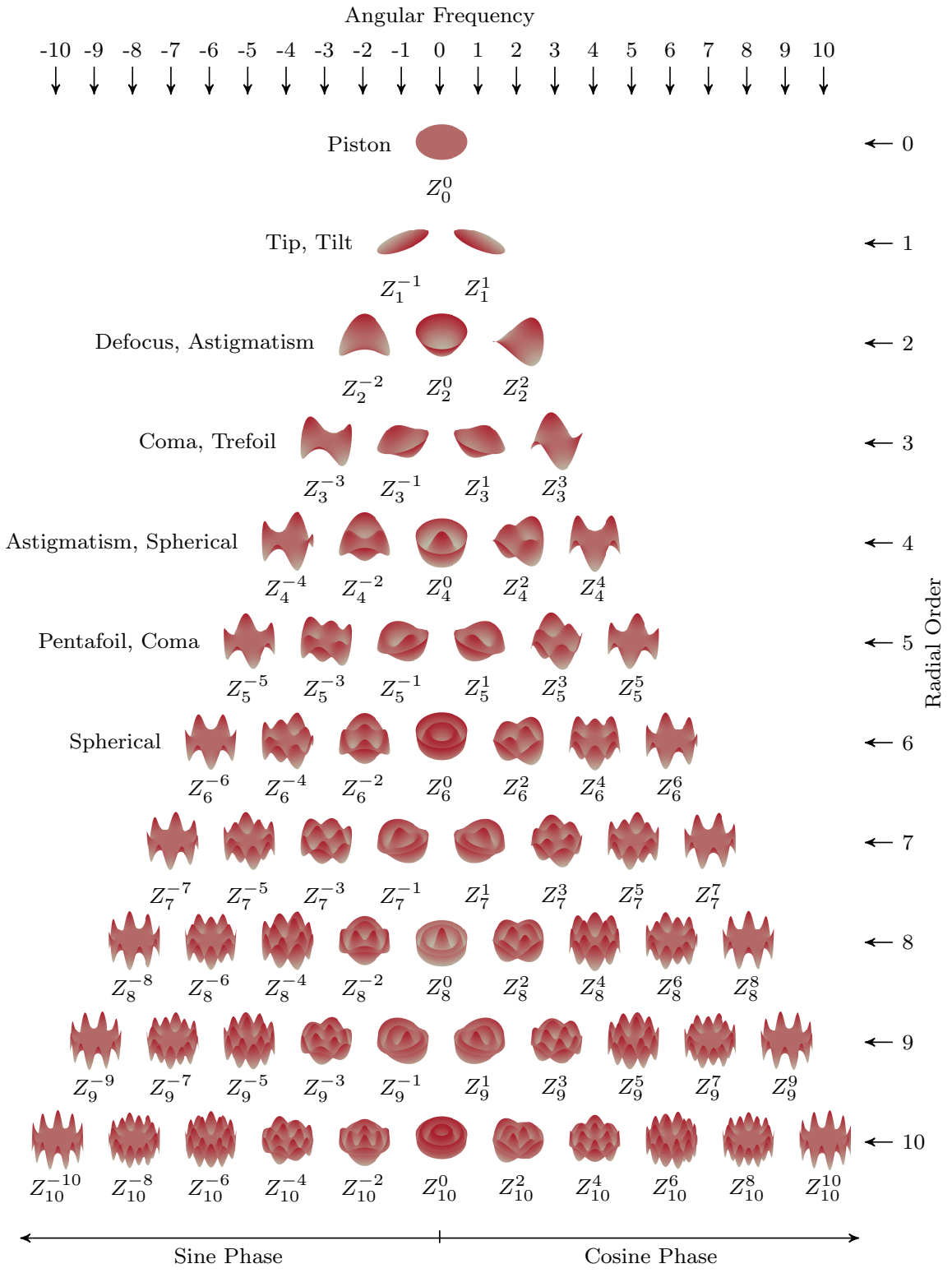


Fig. 2.11 Zernike polynomials

where $I_{x,y}$ is the SLM illumination, $\Phi_{x,y}$ is the aberration phase profile and $H_{\text{uncor},x,y}$ is the uncorrected hologram function.

This can easily be corrected by applying a phase transform of the opposite sign to the hologram

$$E_{u,v} = \mathscr{F}\{I_{x,y}\left[H_{\text{cor},x,y}e^{-2\pi i\Phi_{x,y}}\right]e^{2\pi i\Phi_{x,y}}\} \tag{2.64}$$

$$E_{u,v} = \mathscr{F}\{I_{x,y}H_{\text{cor},x,y}\} \tag{2.65}$$

2.3.4 Field-Dependent Aberration Correction

It is possible to correct for spatially dependent aberrations by using an alternative to the FT. To do so requires an algorithm based on something other than the FT.

2.3.5 Gamma Correction

Displays have a *gamma factor* representing the transfer power function between the input amplitude for a pixel and associated output energy. For an ideal SLM, γ is equal to 2.

$$\hat{V}_{\text{out},x,y} = \left(\hat{V}_{\text{in},x,y}\right)^2 \tag{2.66}$$

where $V_{\text{in},x,y}$ is the SLM input amplitude function and $V_{\text{out},x,y}$ the SLM output energy. In the context of display codecs, compensation is sometimes required.

The imaging camera will also have a gamma factor. This can be determined by illuminating the sensor with a Gaussian beam and fitting the observed power response to the expected normal intensity map.

2.4 Computer-Generated Holograms

A fast Fourier transform (FFT) calculates the discrete Fourier transform (DFT) from an input. As a Fraunhofer hologram can be considered to be a Fourier transform, the inverse FFT (IFFT) of the desired image provides the ideal SLM modulation. If an SLM was able to arbitrarily modulate light in both amplitude and phase then calculating a computer-generated hologram (CGH) would be straight forward. As no such device exists, hologram generation algorithms are required to generate compromise solutions that modulate the light within the constraints of the SLM. For phase modulating SLMs, the problem can be considered a subset of the problem of *phase-retrieval*.

This process is computationally demanding, not least in that scaling with pixel count is a high order problem. A traditional FT has complexity $O(N^4)$ for a square field of dimension N. An FFT has complexity $O(N^2 \log N)$. CGH algorithms exacerbate this, often being $O(N^2)$ themselves. For example, running simulated annealing on every pixel of a field with no further optimisations is $O(N^4 \log N)$. Moving from a 'hd' 1080×1920 to '4k' 2160×3840 display would result in a computation approximately 21 times longer for a display containing only 4 times as many elements. Advanced techniques reduce this scaling but only by a limited amount. Designing processes to run in parallel on modern GPUs brings the problem within the grasp of real-time processing for modern systems.

While a wide array of algorithms are available, the most common of these are summarised in Figure 2.14 on pages 58 - 60. These are expanded on in the sections below.

2.4.1 Performance Metrics

There exist a number of algorithms for finding an appropriate quantised hologram from an idealised hologram. Fundamental to all of these is *quantisation*. Quantisation is the act of adapting a back-projected light field to the display capabilities of the spatial light modulator used.

During hologram generation, tuning is achieved by adjusting the *three freedoms*: *amplitude*, *phase* and *scale*. In a typical application, only a portion of the replay field is of interest, giving amplitude freedom in the other regions. Phase freedom is due to the eye being phase insensitive and scale freedom is provided in applications where image fidelity is more

important than efficiency. Careful adjustment of these factors can cause several orders of magnitude of difference in processing times, Appendix B.

The standard approach to measuring hologram performance is that of mean squared error (MSE). The MSE $E_{\mathrm{MSE},T,R}$ is given as a relation of the target image T and generated replay field R.

$$E_{\mathrm{MSE,PI}}(T,R) = \frac{1}{N_x N_y} \sum_{x=0}^{x=N_x-1} \sum_{y=0}^{y=N_y-1} \left[\left|T_{x,y}\right| - \left|R_{x,y}\right| \right]^2 \tag{2.67}$$

This relationship assumes that the solution is phase insensitive (PI) and that only the spatial intensity profile is of interest. While this is the predominant case discussed in this work, many applications also require phase control and use the following version of MSE.

$$E_{\mathrm{MSE,PS}}(T,R) = \frac{1}{N_x N_y} \sum_{x=0}^{x=N_x-1} \sum_{y=0}^{y=N_y-1} \left|T_{x,y} - R_{x,y}\right|^2 \tag{2.68}$$

The light-matter interaction mechanisms exhibited in additive manufacture, Chapter 3, are purely intensity related. As a result, the image quality metric for all algorithms is assumed to be phase insensitive, unless otherwise stated.

For displays viewed by the human eye, the structural similarity index (SSIM) is more commonly used as it has been shown to more closely correspond to ocular visual quality[48]. Unlike MSE which is determined be comparing the replay and target images, SSIM is determined from moving two 8×8 pixel windows T and R across the target and reconstruction images.

$$E_{\mathrm{SSIM}}(T,R) = \frac{\left(2\mu_T \mu_R + c_1\right)\left(2\sigma_{TR} + c_2\right)}{\left(\mu_T^2 + \mu_R^2 + c_1\right)\left(\sigma_T^2 + \sigma_R^2 + c_2\right)} \tag{2.69}$$

where μ_T and μ_R are the window means; σ_T and σ_R are the window variances; σ_{TR} is the covariance of the two window and c_1 and c_2 are functions of pixel dynamic range, L, where $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$. k_1 and k_2 are usually taken as 0.01 and 0.03 respectively. Pixel dynamic range can be taken as being the total number of possible pixel states.

Fig. 2.12 Standard test images from the USC-SIPI image database[49]. From left to right: Mandrill, Peppers, Man, Camera Man, Aerial, Landscape

2.4.2 Test Images

This thesis carries out tests using the six images selected from the USC-SIPI image database[49] shown in Figure 2.12. The test images used are 512x512 pixels and, unless otherwise stated, the performance metrics are given for this size. The effect of image resolution is discussed later.

There are broadly three categories of algorithms: iterative, search and time-multiplexed. These are discussed in turn.

2.4.3 Iterative Fourier Transform Algorithms

The first family of algorithms are the Iterative Fourier-Transform Algorithms (IFTAs), most famous of which is the Gerchberg-Saxton (GS) algorithms. Appendix B presents a full benchmark and discussion of the parameters effecting GS.

Algorithm 2.1: Gerchberg-Saxton

Input: Target image T and number of iterations N

Output: Output hologram $H \leftarrow H'$

1 Randomise target image phase - Section 2.2.3

$$R'_{u,v} = |T_{u,v}| \angle \text{Rand}[0, 2\pi]$$

for $n \leftarrow 1$ **to** N **do**

2 Back-propagate the target to the diffraction plane:

$$H = \mathscr{F}^{-1}\left\{R'\right\}$$

3 Quantise the resultant hologram:

$$H' = \text{Quantise}\,(H)$$

4 Generate expected image:

$$R = \mathscr{F}\left\{H'\right\}$$

5 Apply target amplitude:

$$R'_{u,v} = |T_{u,v}| \angle R_{u,v}$$

end

Gerchberg-Saxton

GS was presented in 1972[50] and was originally designed for used for phase retrieval[51] before later being applied to generating holograms, Algorithm 2.1[52].

When operating with discrete CGHs, convergence was always the dominant issue and the next few decades saw many techniques for improving this[53]. Today, the major challenge is in choosing the correct compromise between parameter selection, convergence and the chance of being trapped in local minima. Different applications require different choices of merit function. Image generation typically focusses on image quality while beam shaping requires minimisation of power loss. It can be shown that GS minimisesmean squared error (MSE) as well as producing identically independently distributed (i.i.d.) results for i.i.d. randomised phase inputs.

Variants

The IFTA class of algorithms has many other variants that offer performance improvements over GS in specialised cases.

Similar to Gerchberg-Saxton is the Fienup algorithm, later extended by Bauschke, which adjusts the target to reduce reconstruction noise[54].

The phase randomisation step has seen some work modifications including the use of a phase vortex, which smooths the power spectral density (PSD) in a well-understood manner and the optimal rotation angle (ORA) algorithm which restricts the range of randomised values while still reducing edge enhancement[55, 56]. Both methods serve to limit the chance of convergence to local minima but have seen limited adoption due to their relative complexity when compared to the observed performance improvements.

Another family of variants, weighted GS (WGS), includes under- and over-compensation approaches[57]. Here the diffraction side quantisation step where the hologram is constrained to the limits of the SLM is modified to either under- or over-compensate the change by a factor β[58, 59]. This leads to a relationship

$$H' = \text{Quantise}\,(H) = H + \beta\Delta H \tag{2.70}$$

This can improve convergence speed and convergent quality but adds significant overhead in complexity and user expertise.

Many variants on this exist, including *over compensation* where $\beta > 1$, *cooling* where β decreases over time and *variable* where β is dependant on the location in the image. Bearing in mind the three freedoms, page 44, allows for greater freedom in optimising only for the parameters desired.

In addition to weighted constraints is the family of variants which use energy related approaches. For example, some works use techniques for using an initial diffraction field of lower energy than required and scaling the energy over iterations. Replay amplitudes are only constrained to not being greater than the target. This reduces the chance of converging to local minima.

The Liu-Taghizadeh (LT) Algorithm is a well-developed approach that initially restricts the region of interest in the target field

$$H' = \text{Quantise}\,(H) = \begin{cases} H + \Delta H, & \text{where} \quad x, y \in W, \\ H, & \text{otherwise.} \end{cases} \tag{2.71}$$

where W is the region of interest where the constraints are applied and ΔH is the naïve quantization. This reduces the number of degrees of freedom being optimised for at any one point[60]. LT algorithms require experience and time to use but can offer up to 100% improvements in execution time over native GS in specialised cases.

The final main variant group involves replacing the FFT in GS. Fractional Fourier transforms and the gyrator transform have both been used[61, 62]. As with all the algorithms discussed here, these approaches require expertise in use that offsets some of the advantages of their performance. Many variants on this exist, including *over compensation* where $\beta > 1$, *cooling* where β decreases over time and *variable* where β is dependant on the location in the image. Bearing in mind the three freedoms - amplitude, phase and scale - this allows for a cookbook style approach to GS variants[63].

2.4.4 Holographic Search Algorithms

The second family of algorithms are the Holographic Search Algorithms (HSAs). Generally slower than their IFTA counterparts, they are better capable of dealing with ton-smooth problems with many local minima. Two of the most common HSAs, direct search (DS) and simulated annealing (SA) are discussed here. A means of improving their performance is presented in Section 5.2 and a new family of predictive HSAs are developed in Section 5.4.

Direct Search

DS, Algorithm 2.2, is probably the simplest HSA. Simply put, an initial hologram is generated and an error metric $E(T,R)$ defined for determining the error of a hologram relative to the target. Typically this taken as being the mean squared error (MSE), Section 2.4.1[64, 65].

Algorithm 2.2: Direct Search

Input: Target image T, number of iterations N and error function $E(T,R)$

Output: Output hologram $H \leftarrow H'$

1 Randomise target image phase - Section 2.2.3

$$R_{u,v} = |T_{u,v}| \angle \text{Rand}[0, 2\pi]$$

2 Back-propagate the target to the diffraction plane:

$$H = \mathscr{F}^{-1}\{R\}$$

3 Quantise the resultant hologram:

$$H' = \text{Quantise}(H)$$

4 Generate initial replay field:

$$R = \mathscr{F}\left\{H'\right\}$$

5 Generate initial error:

$$E = \text{Error}(T,R)$$

for $n \leftarrow 1$ **to** N **do**

6 Modify a random pixel's value to give $H' \leftarrow H$

7 Generate expected image:

$$R = \mathscr{F}\left\{H'\right\}$$

8 Generate expected error:

$$E'_n = \text{Error}(T,R)$$

if $E' > E$ **then**

9 Undo the pixel flip by resetting value of $H' \leftarrow H$

else

10 Advance $H_{n+1} \leftarrow H', \quad E_{n+1} \leftarrow E'$

end

end

Many variants of DS exist, primarily aimed at improving convergence and reducing the chance of becoming trapped in local minima. Improvements typically hinge around using multiple pixels and improved pixel selection algorithms[66, 67]. Computation times can be further improved by using lookup tables (LUTs) of the effects of a single pixel flip and DC balance issues can be mitigated by swapping pixel values to ensure a constant total switched state[68].

A significant subcategory of DS algorithms employ *simulated annealing* to mitigate local minima which is discussed in Section 11. Allowing spatial and phase freedom, page 44, has been shown to significantly improve convergence[69]. Alternative formulations of the error function allow for optimisation for different variables such as efficiency.

For the case of switching a single pixel, the FFT step shown in Algorithm 2.2 can be replaced with an updated state. FFT operations run in $O(N_xN_y\log(N_xN_y))$ time whereas the following update step can be run in $O(N_xN_y)$ time. For changing a pixel $H_{m,n}$, the new replay field value $R_{u,v}$ is given by

$$\Delta R_{u,v} = \frac{1}{\sqrt{N_xN_y}}\Delta H_{x,y}e^{\left[-2\pi i\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)\right]} \tag{2.72}$$

Simulated Annealing

Simulated annealing (SA) Algorithm is a probabilistic approach to discrete domain problems with unpredictable local-minima, so called as it approximates the recrystallisation process during metal annealing[70]. First used in the 1980s, SA has become a major part of the algorithm designers handbook[71].

SA avoids a steepest descent hillwalking approach, instead choosing its next iteration probabilistically. Over time the *temperature* is decreased with the probability of iteration acceptance depending on a probability function, P. While there are many variants, an example is given in Algorithm 2.3.

After a single pixel is modified, the error E_{n+1} of the resulting target is calculated and compared with the original error E_n to give ΔE. If $\Delta E > 0$ then the chance of acceptance is chosen using a modified Boltzmann function.

Algorithm 2.3: Simulated Annealing

Input: Target T, number of iterations N, error function $E(T,R)$, initial temperature t_0

Output: Output hologram $H \leftarrow H'$

1 Randomise target image phase - Section 2.2.3

$$R_{u,v} = |T_{u,v}| \angle \text{Rand}[0, 2\pi]$$

2 Back-propagate the target to the diffraction plane:

$$H = \mathscr{F}^{-1}\{R\}$$

3 Quantise the resultant hologram:

$$H' = \text{Quantise}(H)$$

4 Generate initial replay field:

$$R = \mathscr{F}\left\{H'\right\}$$

5 Generate initial error:

$$E_0 = \text{Error}(T,R)$$

for $n \leftarrow 1$ **to** N **do**

6 Update temperature: $T = t_{\text{coeff}} e^{-t_0 \frac{n}{N}}$

7 Modify a random pixel's value to give $H' \leftarrow H$

8 Generate expected image:

$$R = \mathscr{F}\left\{H'\right\}$$

9 Generate expected error:

$$E' = \text{Error}(T,R)$$

if $E' - E > 0$ *and* $e^{\frac{E'-E}{t}} < Rand[0,1]$ **then**

10 Undo the pixel flip by resetting value of $H' \leftarrow H$

else

11 Advance $H \leftarrow H', \quad E \leftarrow E'$

end

end

Algorithm 2.4: One-Step Phase-Retrieval

Input: Target image T and number of sub-frames N

Output: Output holograms $H[1..N] \leftarrow H'[1..N]$

for $n \leftarrow 1$ **to** N **do**

1 Randomise target image phase - Section 2.2.3

$$R'_{u,v} = |T_{u,v}| \angle \text{Rand}[0, 2\pi]$$

2 Back-propagate the target to the diffraction plane:

$$H = \mathscr{F}^{-1}\left\{R'\right\}$$

3 Quantise and output the resultant hologram:

$$H'_n = \text{Quantise}\,(H)$$

end

$$P(\Delta E) = e^{\frac{-\Delta E}{t}} \tag{2.73}$$

$$t = t_{\text{coeff}} e^{-t_0 \frac{n}{N}} \tag{2.74}$$

The range of applications for SA is extensive within holography and there are as many uses as there are researchers[72]. As with DS, SA can be improved by using an update function such as (2.72).

2.4.5 Time-Multiplexed Algorithms

The final family of algorithms are the time-multiplexed algorithms (TMAs). The most common of these, one-step phase-retrieval (OSPR) and its variant adaptive OSPR (AdOSPR) are introduced here. The issue of OSPR performance is discussed in Appendix A. A new algorithm that improves on both OSPR and adaptive OSPR is presented in Section 5.3.

One-Step Phase-Retrieval

The OSPR algorithm shown in Algorithm 2.4 was developed for real-time hologram generation for projectors and relies on averaging many low-quality holograms[73]. This approach relies on time averaging with fast switching SLMs displaying many independently distributed sub-frames generated from a phase randomised starting point with the error variance being time averaged.

The normalised human eye cone cells impulse response drops by 30% in the first 20 ms after an event[74]. This is often modelled as the eye failing to distinguish features faster than 60 frames per second (FPS). Showing many subframes in that interval is imperceptible to the human eye and forms the basis of OSPR algorithms where time averaging is used to to produce a visually high quality replay field from a a large number of low quality holograms.

The variance of the time averaged light field follows a reciprocal relationship with the number of sub-frames N.

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}\frac{\sigma_n^2}{N} \tag{2.75}$$

Adaptive One-Step Phase-Retrieval

A variant of OSPR, adaptive OSPR is shown in Algorithm 2.5. In this system, each target frame is adjusted in relation to the frames before. This adds significant additional overhead but also significantly improves noise averaging[43]. This leads to a square law on the error.

2.4.6 Other Algorithmic Approaches

Neural networks[75] and Genetic algorithms[76, 77] have been investigated but are not currently used in real-time applications due to the high-performance cost. Hybrid algorithms combining IFTA and SA are also used[78].

Also of interest are error diffusion (ED) or error propagation (EP) approaches where a pixel is quantised and then the error propagated across the rest of the remaining pixels by adjusting the target image after each quantisation step[79]. Physical systems have been developed

Algorithm 2.5: Adaptive One-Step Phase-Retrieval

Input: Target image T and number of sub-frames N

Data: Total visual field generated so far F

Output: Output holograms $H[1..N] \leftarrow H'[1..N]$

for $n \leftarrow 1$ **to** N **do**

1 Randomise target image phase - Section 2.2.3

$$R'_{u,v} = |T_{u,v}| \angle \text{Rand}[0, 2\pi]$$

2 Back-propagate the target to the diffraction plane:

$$H = \mathscr{F}^{-1}\left\{R'\right\}$$

3 Quantise and output the resultant hologram:

$$H'_n = \text{Quantise}\,(H)$$

4 Add expected image to accumulated light field:

$$F \leftarrow= F + \mathscr{F}\left\{H'\right\}$$

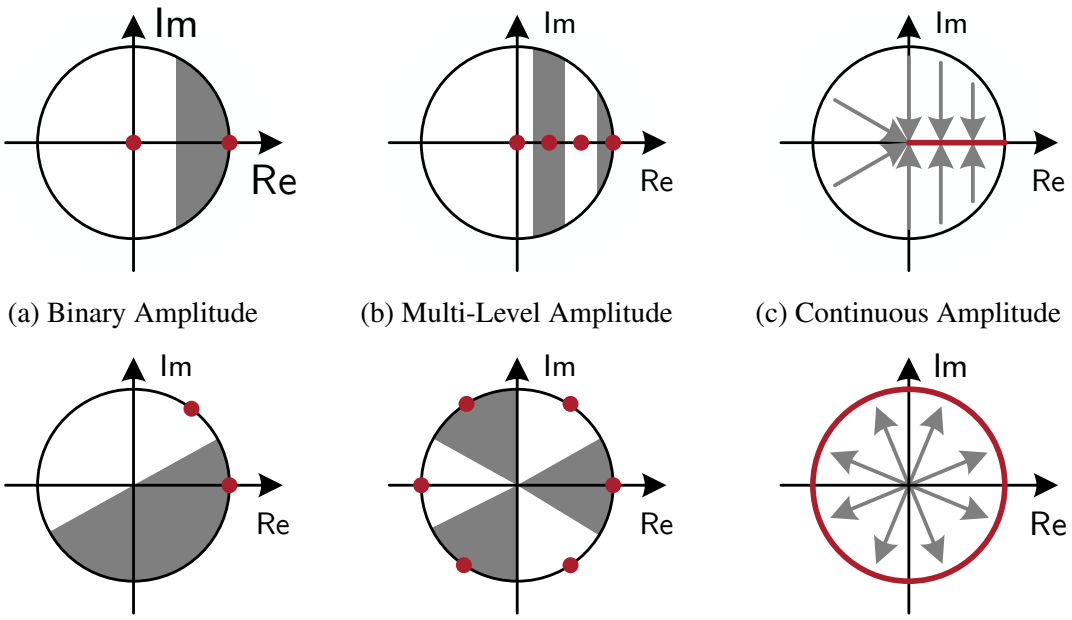
5 Generate compensation for next frame:

$$T \leftarrow T + \frac{nT - F}{N - n}$$

end

where the mathematics for hologram decomposition is emulated on a physical system with a second SLM[80].

More recently, the wider field of phase retrieval has seen the rise of new methods including semidefinite programming (SDP), two-stage sparse phase retrieval (TSPR) and Wirtinger flow (WF)[81–85]. All three techniques involve embedding the two dimensional image matrix into a higher order system that can be treated as linear and convex[86]. While these approaches offer better results even in piecewise and noisy systems, they are fundamentally unsuited to display holography as iteration times are at least of $O(N^4)$.



(a) Binary Amplitude (b) Multi-Level Amplitude (c) Continuous Amplitude

(d) Binary Phase (e) Multi-Level Phase (f) Continuous Phase

Fig. 2.13 Quantisation of common modulation schemes. c.f. Figure 1.6

2.4.7 Quantisation Techniques and DC Balance

Typically, the quantisation steps in the algorithms discussed in the sections above are as simple as rounding to the nearest available step value. E.g. for a normalised binary amplitude system

$$H'_{x,y} = \begin{cases} 1, & H_{x,y} > \frac{1}{2}, \\ 0, & \text{otherwise} \end{cases} \tag{2.76}$$

The *nearest-level* or *nearest-neighbour* approach for each of the six basic modulation schemes in Figure 1.6 is shown in Figure 2.13. A number of other techniques have been developed and this is discussed further in Section 5.5.

Quantisation can lead to significant DC balance issues which can destroy the SLM due to electrostatic charging so various weighted averaging techniques have been developed. A typical example might be the following which examines the surrounding 8 cells[87].

Table 2.3 Hardware alternatives for computer-generated holograms. Adapted from[89]

	Core Count	Development Period	Clock Frequency	Power Consumption	Generation Time
CPU	Low	Short	High	Average	High
FPGA	High	Long	Low	Low	Average
GPU	High	Average	High	High	Low

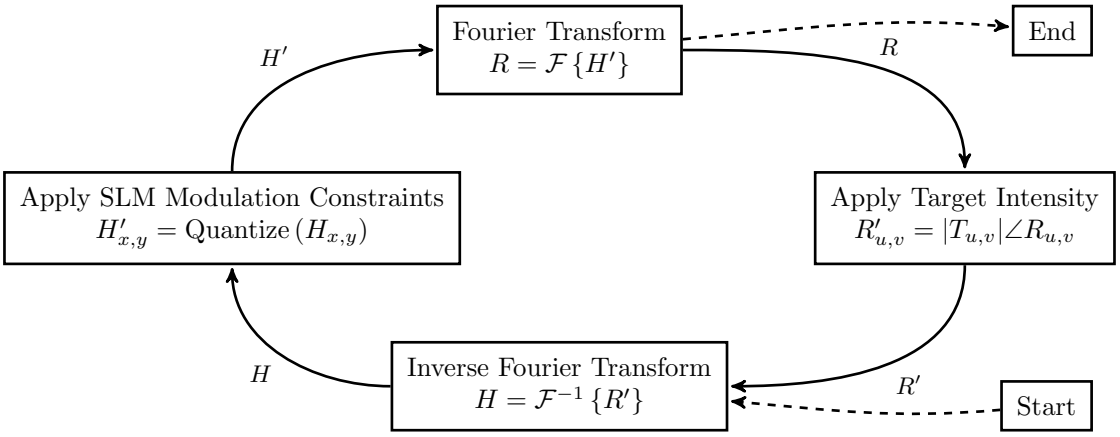
$$H'_{x,y} = \begin{cases} 1, & H_{x,y} > \frac{\sum_{n=x-1}^{x+1} \sum_{m=y-1}^{y+1} H_{m,n}}{9}, \\ 0, & \text{otherwise} \end{cases} \tag{2.77}$$

This smoothing also helps reduce the edge enhancement effects discussed in Section 2.51.

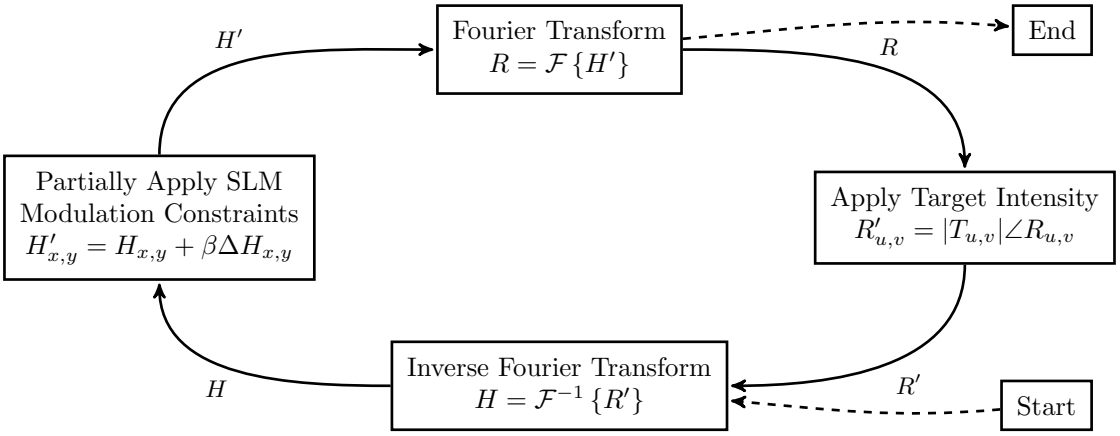
There are a number of other quantisation techniques that can be used to improve the result. These can then be integrated with existing approaches such as Gerchberg-Saxton. *Error diffusion* is used where each pixel is quantised and then the errors of the remaining pixels updated to influence future pixel quantisations[88] Their use is beyond the scope of this work, however.

2.4.8 Limitations

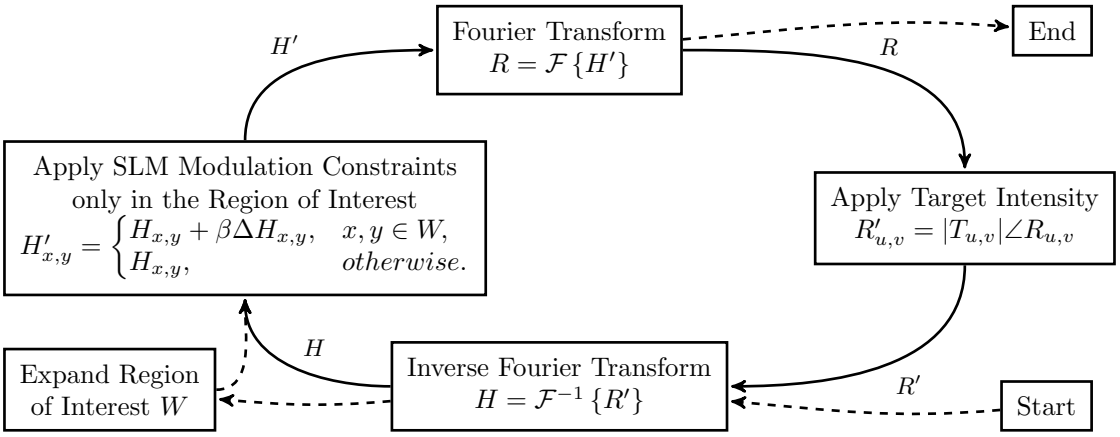
Expensive display hardware and the high computational load mean that holographic projectors have failed to take off [90, 91]. High power projectors are typically of the order of £10,000s which prohibits them from many applications. In the case of AM where machines cost £100,000s, areal projection systems offer the potential to reduce costs as well as increase performance. Table 2.3 provides a comparison of the three different hardware configurations used in hologram generation. While field programmable gate arrays (FPGAs) will be discussed later, GPUs will be the primary focus of this work due to their flexibility, low generation latency and shorter development time.



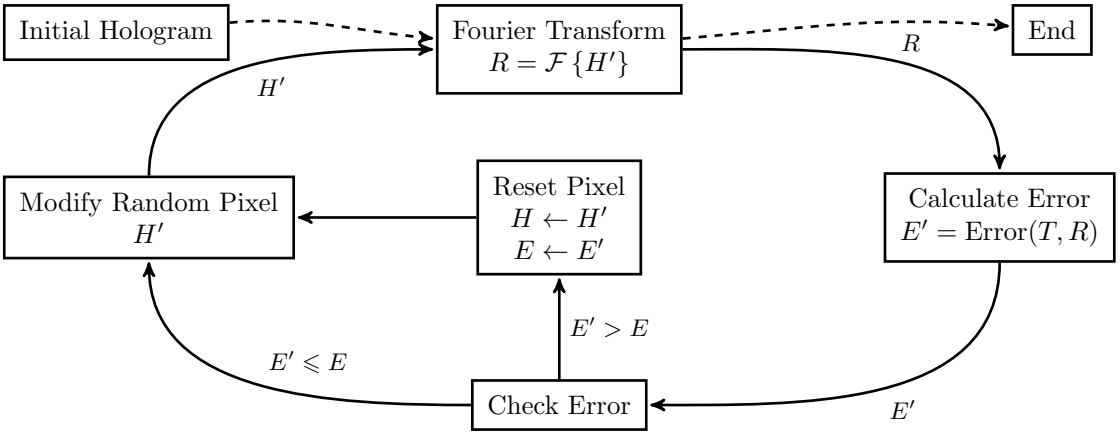
(a) Gerchberg-Saxton



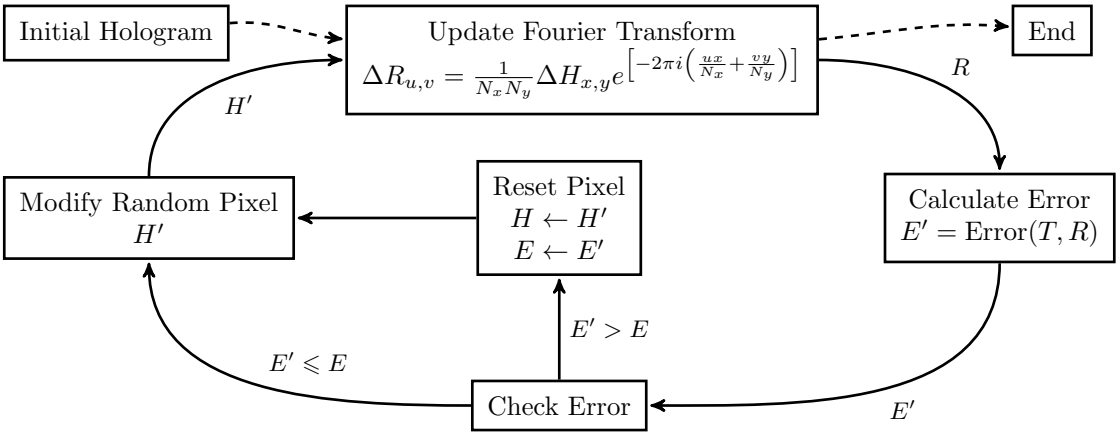
(b) Weighted Gerchberg-Saxton. Compare (a)



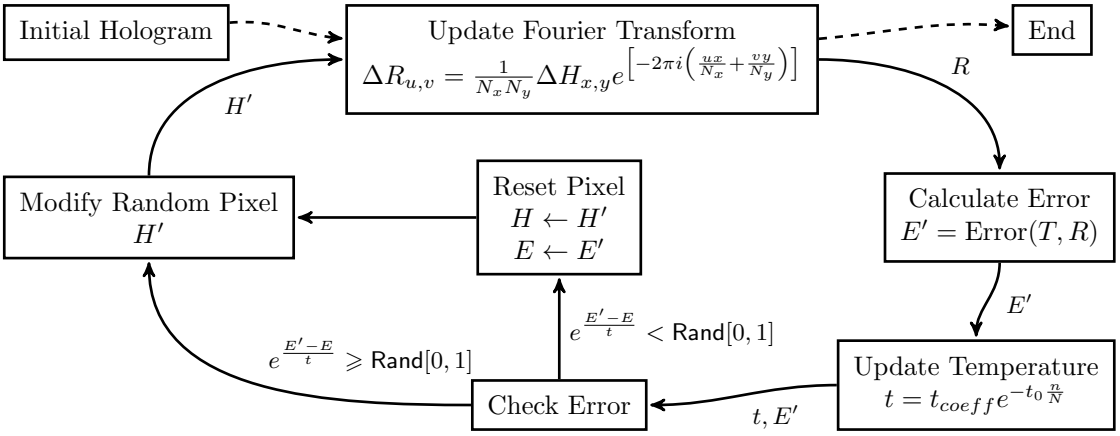
(c) Liu-Taghizadeh. Compare (a)



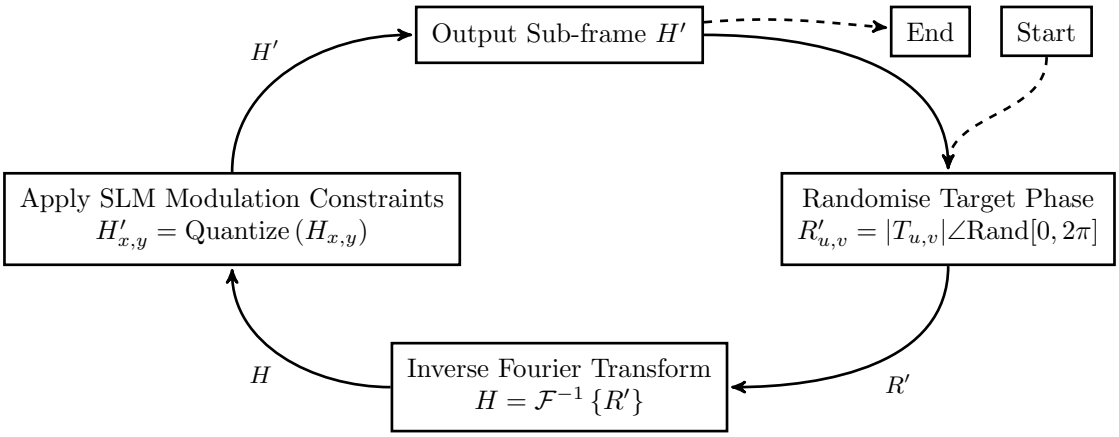
(d) Naive Direct Search



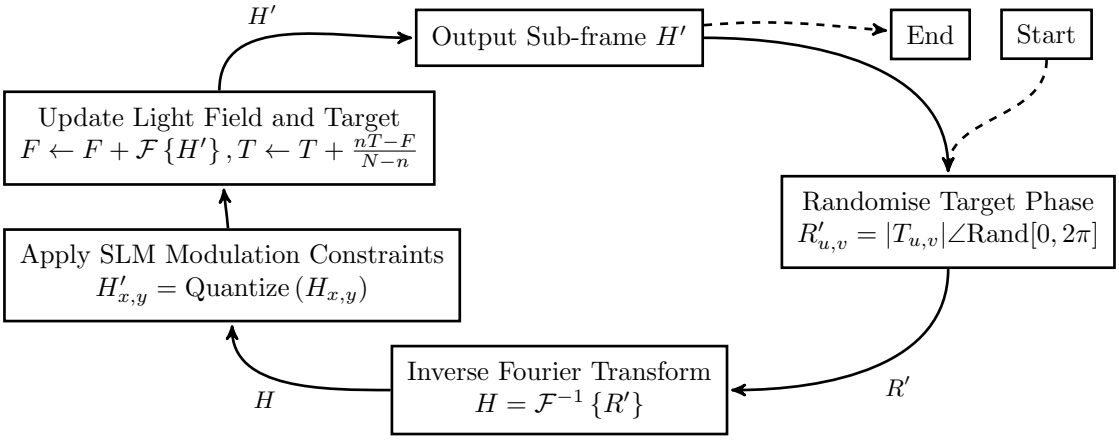
(e) Direct Search. Compare (d)



(f) Simulated Annealing. Compare (d)



(g) One-Step Phase-Retrieval



(h) Adaptive One-Step Phase-Retrieval. Compare (g)

Fig. 2.14 Common computer-generated hologram algorithms

Chapter 3

Additive Manufacture Modelling

Chapter 3

Additive Manufacture Modelling

3.1 Modelling

This chapter sets out to introduce the different techniques for modelling additive manufacture (AM) interactions as well as presenting a more detailed understanding of the underlying processes. The literature in this field is enormous. Six recent review papers of different aspects listed over a thousand independent references between them[92–98]. The salient details from the field relevant to this work are summarised in the sections below.

3.1.1 Beam Profile

Laser spatial energy distributions, Figure 3.1, are described by transverse electromagnetic (TEM) modes with TEM_{00} representing the idealised Gaussian beam given by[99]

$$I(r) = I_0 e^{-r^2} \tag{3.1}$$

where I represents the intensity at distance r from the centre of the beam and I_0 is the average beam intensity. The beam edge is conventionally demarcated as either being the point where the intensity drops to $\frac{1}{e^2}$ (D_{86}) or as the full-width half-maximum (FWHM).

With a beam quality factor (BQF) of M^2, focal length f and wavelength λ the minimum focal spot diameter (FSD) D_0 is

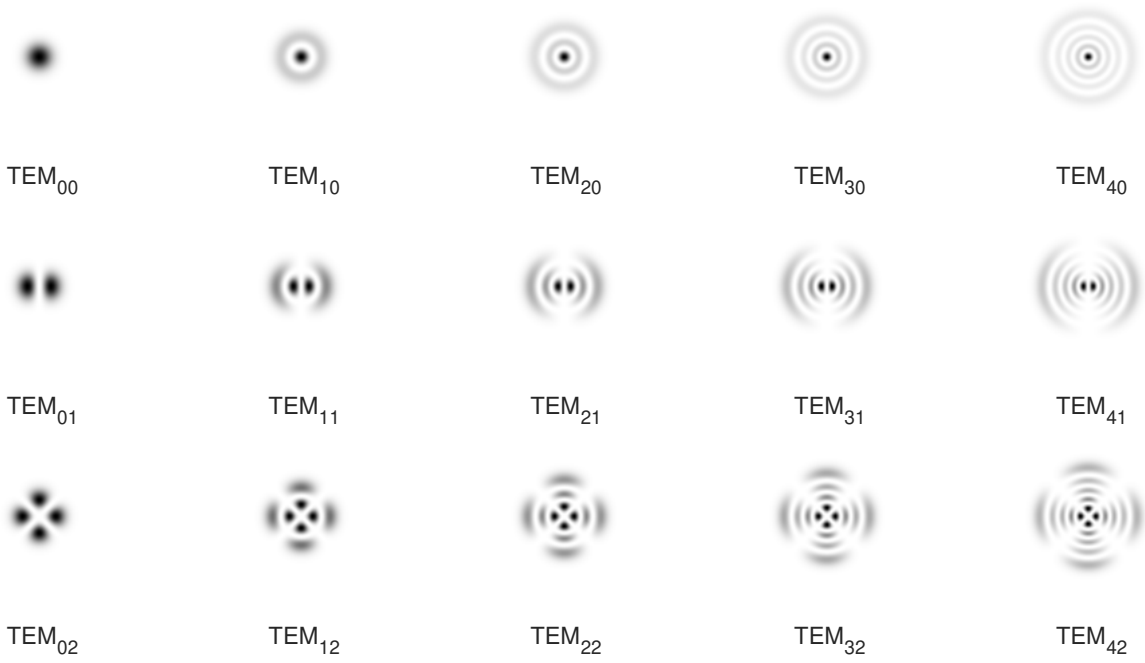


Fig. 3.1 Transverse electromagnetic modes

$$D_0 = \frac{\lambda f M^2}{\pi D} \tag{3.2}$$

According to Figure 3.2, at distance z axially from focus with *propagation constant* $k = \frac{2\pi}{\lambda}$

$$D(z) = D_0 \left[1 + \left(\frac{2zM^2}{kD_{min}^2} \right)^2 \right]^{\frac{1}{2}} \tag{3.3}$$

The depth of focus is given by the Rayleigh range length and is equal to[100].

$$z_R = \frac{\pi w_0^2}{\lambda} \tag{3.4}$$

The depth of focus or confocal parameter is then $b = 2z_R$

Beam profiles are usually assumed to follow an axisymmetric Gaussian profile with radial and axial coordinates (r,z)[99].

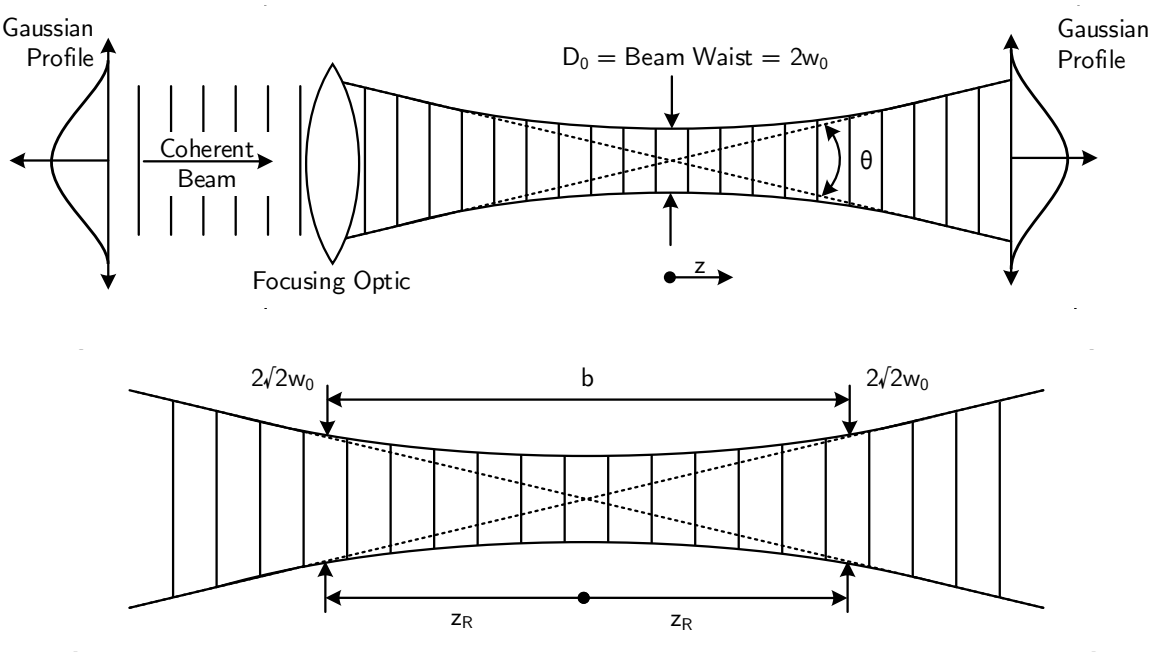


Fig. 3.2 Beam profile

$$I(r,z) = \frac{2P}{\pi w(z)^2} e^{\left(-2\frac{r^2}{w(z)^2}\right)} \tag{3.5}$$

where P is the total *output* power of the heat source, $w(z)$ is the radius at distance z from the waist and $I(r,z)$ the beam intensity at a point.

The following applies

$$\int_0^{2\pi} \int_0^{\infty} rI(r,z)drd\theta = P \tag{3.6}$$

The pre-eminent factor in traditional PBF is the power density of the laser[101]. The *peakedness* of the Gaussian beam is termed the distribution factor f, Figure 3.3, and leads to this alternative formulation of (3.5) from the heat source radius.

$$I(w,z) = \frac{fP}{\pi w(z)^2} e^{\left(-f\frac{w_b^2}{w(z)^2}\right)} \tag{3.7}$$

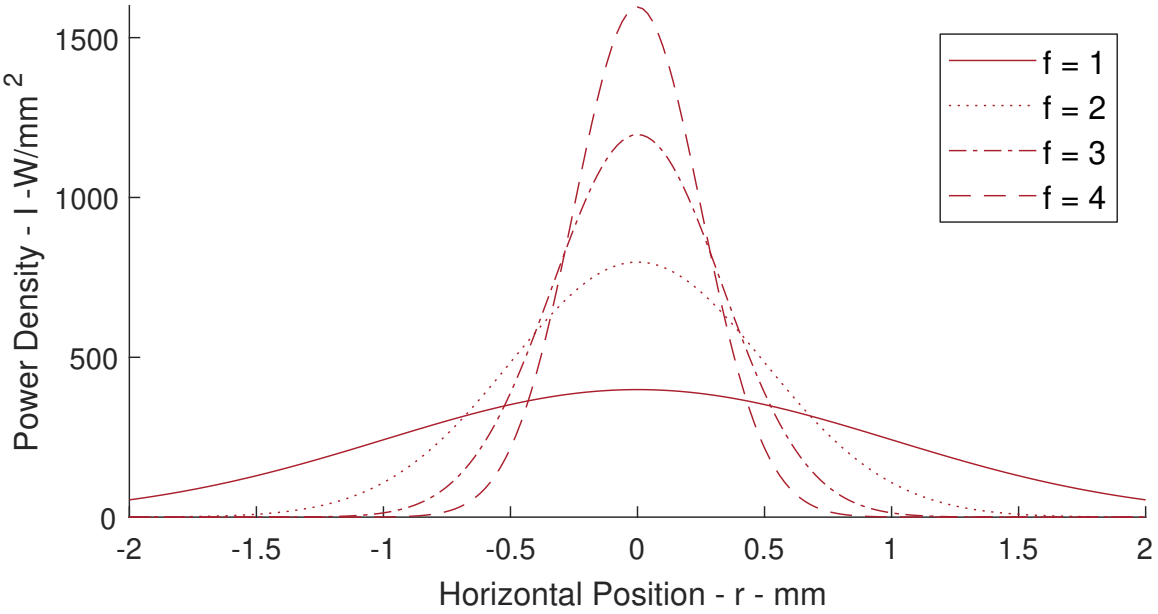


Fig. 3.3 Power density distribution from power distribution factor

Typical PBF laser spot profiles are in the range of $50-100\mu$m. power density distribution (PDD) measurements for high power lasers is typically via a rotating wire and measurement at points along the beam allowing for determination of the beam divergence and M^2 factor[102]. Electron beams can be measured with a faraday cup (FC), diffraction slits or pinholes[103, 104]. Laser profiles are often shaped using fixed optical systems[29].

3.1.2 Heat Source Modelling

3.1.3 Conservation Equations

For the heat source in (3.7), a fraction η_i is absorbed in the topmost layer, giving the following relationship for absorbed power P_a from layer thickness t_i[105].

$$I(w,z) = \eta_l \frac{fP}{\pi w(z)^2 t_l} e^{\left(-f\frac{w_b^2}{w(z)^2}\right)} \tag{3.8}$$

Lost energy $1-\eta_l$ is due to surface reflectance, absorption by the inert gas, absorption by lower layers and melt pool spatter from boiling and electrostatic forces, Figure 3.4. This absorption coefficient is higher than the Fresnel coefficient of the molten surface[106, 107].

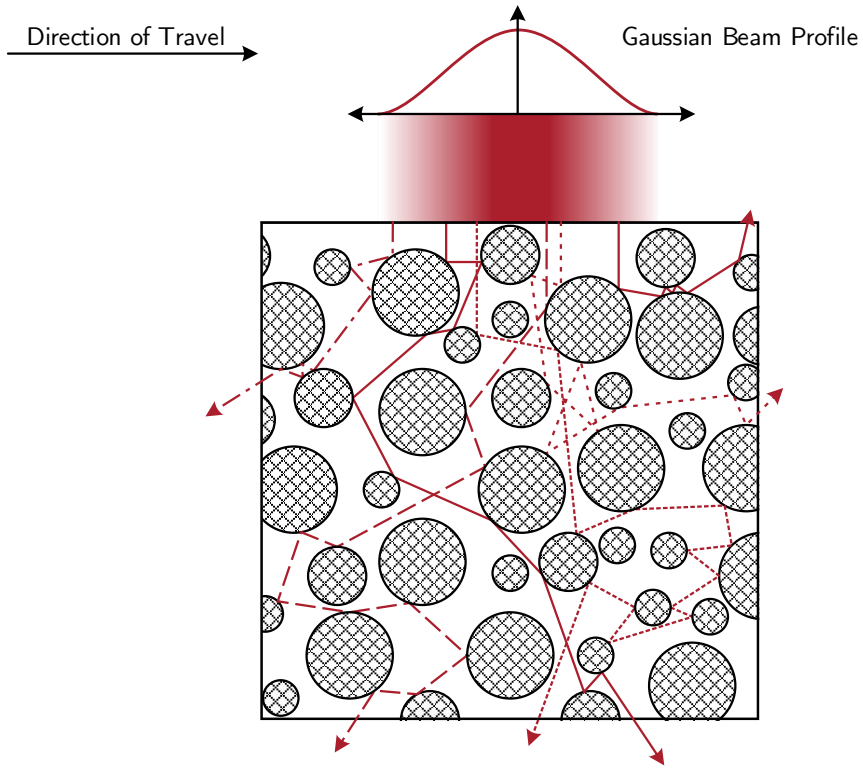


Fig. 3.4 Incident light scattering

This section aims to present the mathematics behind modelling a simple L-PBF process.

Underneath a moving laser spot, massive thermal cycling occurs. This causes inhomogeneous location dependent microstructure and properties. Temperature measurements can only be made on the surface and 3d models are used for determining the interior distortions and residual stresses. It is commonly recognised that AM has more in common with welding than with casting and the majority of relevant mathematics stems from there.

The small feature sizes, long time frames and complex physical phenomena make PBF an extremely difficult process to model[108]. For example, it is possible to estimate that using a standard finite element package for modelling printing a metallic volume of $1m^3$ on a high-end machine wouldn't have finished before the heat death of the universe. Massive improvements in performance can be achieved by the use of multi-scalar modelling. Common simplifications include ignoring Marangoni convection (MC), vaporisation and gas interface behaviour as well as modelling the heat input as a 2D ellipse rather than the measured PDD[109]. Unfortunately, ignoring melt pool flows cause significant error and have been shown to significantly impact porosity, surface structure, spatter and denudation[110–112].

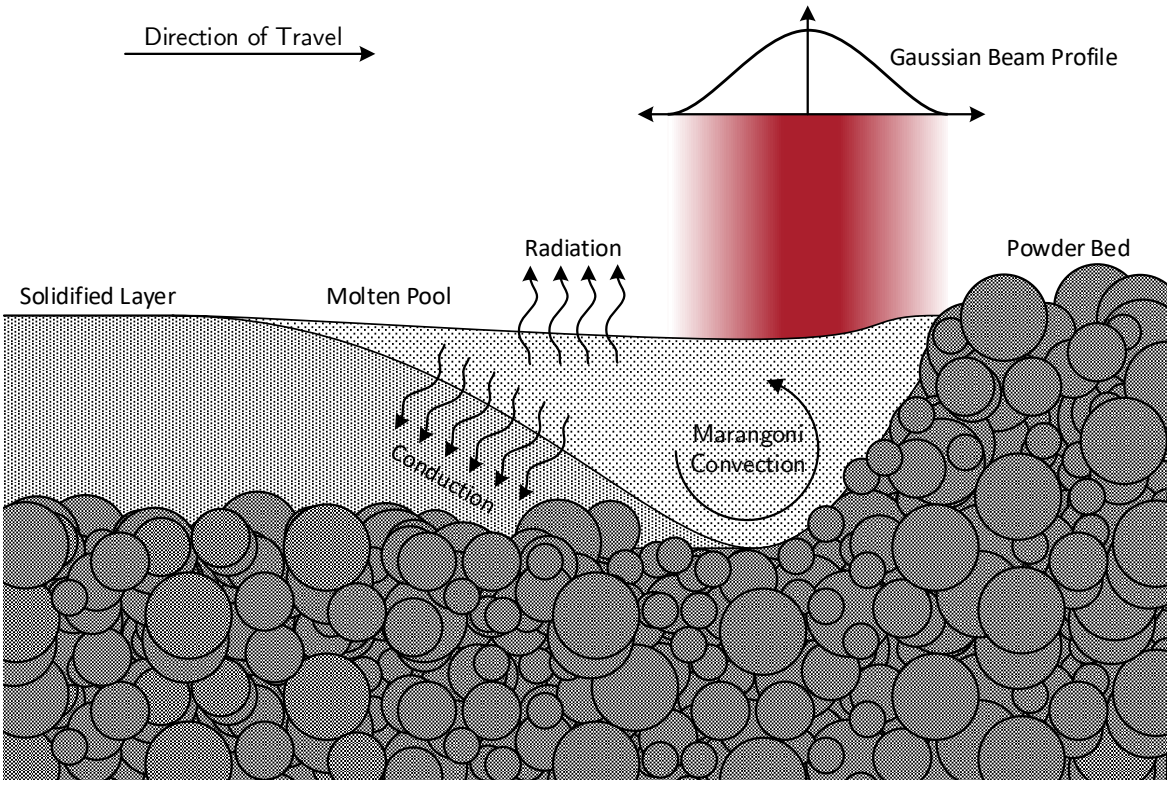


Fig. 3.5 Meltpool behaviour

The temperature behaviour is modelled by solving the mass, momentum and energy equations

$$\frac{\partial(\rho u_i)}{\partial x_i} = 0 \tag{3.9}$$

$$\frac{\partial(\rho u_j)}{\partial t} + \frac{\partial(\rho u_i u_j)}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\mu \frac{\partial u_i}{\partial x_i}\right) + S_j \tag{3.10}$$

$$\rho \frac{\partial h}{\partial t} + \frac{\partial(\rho u_i h)}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{k}{C_p}\frac{\partial h}{\partial x_i}\right) - \rho \frac{\partial \Delta H}{\partial t} - \rho \frac{\partial(u_i \Delta H)}{\partial x_i} \tag{3.11}$$

where u_i and u_j are the i and j velocity components, x_i and x_j the i and j direction components, k is the thermal conductivity, C_p the specific heat, h the sensible heat and ΔH the latent heat component. The term S_j is used to cover buoyancy and elecromagnetic forces and can often be treated as 0 in PBF processes[113]. The modelling of melt pool rheology can only be indicitive as there are limited mechanisms available for measuring flows in-situ[114].

Table 3.1 Compositions of common AM alloys. Updated from[1].

Alloy	Ti	Al	V	Fe	Ni	Cr	Mn	Mg	Si	Mo
SS 316		0.005		Rem.	8.26	17.2	1.56		0.33	
Ti-6Al-4V	Rem.	6.28	3.97	0.052						
IN 718	1.02	0.50		Rem.	53.4	18.8	0.07		0.12	2.99
800 H	0.35	0.25		Rem.	31.0	20.6	0.85		0.32	
H 13			1.20	Rem.		5.50	0.60		1.25	1.75
AA 6061	0.15	Rem.		0.7			0.15	1.2	0.8	

3.1.4 Boundary Conditions

The surface tension gradient on the surface of the melt pool is the primary driver of mass flow and is termed Marangoni flow. This is given by

$$\tau_M = \frac{d\gamma}{dT}\frac{dT}{dr} \tag{3.12}$$

$$= -\mu\frac{du_i}{dx_k} \tag{3.13}$$

where γ is the surface tension, T the temperature, τ_M the Marangoni stress, r the radial distance from heat source axis and x_k the vertical distance from the surface[115].

The alloy temperatures and composition, Table 3.1, dominate the surface tension behaviour[116]. For example, for most unalloyed metals, $\frac{d\gamma}{dT} < 0$ and the molten liquid flows towards the meltpool periphery causing a large wide molten pool. When active elements are added, $\frac{d\gamma}{dT}$ is often positive and the pool tends to draw inwards, giving a narrow and deep pool with resultant heat input deeper within the material[117].

If the flow velocity is taken as zero both vertically and at the interface then

$$-k\frac{\partial T}{\partial z} = \sigma_{SB}\varepsilon\left(T^4 - T_a^4\right) + h_c\left(T - T_a\right) \tag{3.14}$$

where h_c is the convective heat transfer coefficient, T_a the ambient temperature, ε the emissivity and σ_{SB} is the Stefan-Boltzmann constant. The heat transfer coefficient must be

determined experimentally[118]. The thermo-physical properties of key alloys are presented in Table 3.2[119, 120].

A number of approaches have been taken to solving these equations:[121]

- **Analytical Method** - With sufficient simplifications the equations can be solved analytically for simple cases. While this is less computationally expensive it is known to be highly inaccurate[122–124].
- **Finite Element Analysis (FEA)** - FEA packages are widely available and stable however solution times are very high. FEA doesn't properly account for fluid flow and therefore tends to overestimate peak temperatures[125–127].
- **Finite Difference Method (FDM)** - Similar to FEM but is better suited to modelling fluid flow behaviour. Requires a finer grid size and is less capable of modelling material deformations[128, 129].
- **Level Set Method (LSM)** - Model based on the surface of the molten pool. LSM agrees well with experiments but highly computationally expensive[130, 131].
- **Volume of Fluid (VOF) using FDM** - Similar to LSM but based on a volumetric algorithm[132].
- **Lattice Boltzmann Method (LBM) and Arbitrary Lagrangian-Eulerian (ALE)** - Simulates a reduced molecular behaviour based on a cellular automation approach. Very good performance but extremely computationally intensive[133, 134].

Probably the dominant strategy at present is FEA. To varying degrees these incorporate the multi-scalar approaches required for AM and can offer significant performance benefits. The addition of material during solution offers a particular challenge for traditional systems.

Table 3.2 Thermo-physical properties of common AM alloys[119, 120]. Updated from[1].

Alloy	Liquidus Temperature K	Solidus Temperature K	Density kg/m^3	Viscosity $kg/(m \cdot s)$	$d\gamma/dT$ $N/(m \cdot K)$	Thermal Conductivity* $W/(m \cdot K)$	Specific Heat* $J/(kg \cdot K)$
SS316	1733	1693	7800	7×10^{-3}	-0.40×10^{-3}	$A = 11.82$	$A = 330.9$
						$B = 0.0106$	$B = 0.563$
							$C = -4.015 \times 10^{-4}$
							$D = 9.465 \times 10^{-8}$
Ti-6Al-4V	1928	1878	4000	4×10^{-3}	-0.26×10^{-3}	$A = 1.57$	$A = 492.4$
						$B = 1.6 \times 10^{-2}$	$B = 0.025$
						$C = -1.9 \times 10^{-6}$	$C = -4.18 \times 10^{-6}$
IN 718	1609	1533	8100	5×10^{-3}	-0.37×10^{-3}	$A = 0.56$	$A = 360.4$
						$B = 2.9 \times 10^{-2}$	$B = 0.026$
						$C = -7.0 \times 10^{-6}$	$C = -4.0 \times 10^{-6}$
H13 Steel	1725	1585	7900	7×10^{-3}	-0.43×10^{-3}	$A = 1829$	$A = 341.9$
						$B = 7.5 \times 10^{3}$	$B = 0.601$
							$C = -4.04 \times 10^{-6}$
AA 6061	925	855	2700			$A = 2.52$	$A = 929.0$
						$B = 0.4 \times 10^{-2}$	$B = -0.627$
						$C = -7.36 \times 10^{-6}$	$C = -1.48 \times 10^{-3}$

*Given in the form $A + BT + CT^2 + DT^3$

3.2 Material Interactions

3.2.1 Introduction

Light-material interactions (LMIs) are fundamental to the additive manufacture process

All physical systems deviate from the ideal. The significant advantage of SLMs is that, where traditionally a complex optical chain would be used for corrections, all corrections can be included in the SLM. This is a significant advantage for high-power AM in a dirty environment as it reduces failure points, energy absorption and total cost for large optics.

3.2.2 Powders

Manufacture of PBF powders is a significant challenge with a number of the best-performing systems being hidden by propriety systems[135]. Powder is probably the biggest unknown and cause of many issues[136]. Not all powders are created equal and measurement of powder properties is involved[137, 138].

Shape, size, morphology, flowability, composition and porosity are important. Typical powders are 10-100 μms in diameter. computed tomography (CT) and scanning electron microscopes (SEMs) can be used for determining surface morphology while energy dispersive x-ray spectroscopy (EDS), laser-induced breakdown spectroscopy (LIBS) and x-ray photoelectron spectroscopy (XPS) can be used for determining the composition. Graduated fine sieves are used to determine the range of powder sizes while flowability is determined by using a Hall Meter[139]. indexScanning Electron Microscope

There are four primary means of powder production:

- **Gas Atomisation (GA)** - Atomisation occurs by spraying molten alloy through a high-pressure gas jet[140, 141].
- **Water Atomisation (WA)** - Atomisation occurs in a similar manner to GA but solidification occurs in a high-speed water jet[142].
- **Rotary Atomisation (RA)** - Atomisation occurs by molten alloy being pulled off of a rotary disk through centripedal force[143].

- **Plasma Rotating Electrode Process (PREP)** - Atomisation occurs by a rod with the end heated by a plasma jet[144].

Powders should ideally be spherical and of uniform size and good powder properties are essential for high-quality surface finishes and low porosity. Many practical and theoretical studies have been carried out[145, 146].

3.2.3 Thermal Behaviour

In-Situ Measurement

The transient and localised nature of additive melt pools mean that temperature measurement is difficult[147]. The centre of the meltpool can reach temperatures several hundred degrees over the liquidus temperature of the alloy and achieve boiling. Thermocouples are the most common approach due to their high-temperature range but they are unable to measure more than locally and a very low thermal inertia is required[148]. Infrared thermography can be used to give a less accurate temperature reading of the entire surface but is not capable of detecting temperatures below the surface[149].

Cooling

The product of the *solidification growth rate* (SGR) and the *cooling temperature gradient* (CTG) is a good determiner for the morphology of the final microstructure. For casting, this is typically in the range $[10^0, 10^2]\mathrm{K/s}$ and doesn't go above $[10^2, 10^3]\mathrm{K/s}$ for traditional welding processes. For PBF additive processes this ranges of $[5 \times 10^4, 6 \times 10^6]\mathrm{K/s}$ have been reported[150, 151]. While useful, this value is only indicative as it fails to cover the many variations in the cooling phenomena. Two cooling rates, in particular, are worthy of notice. The liquidus-solidus rate where dendritic structures are determined and the 800-500° rate in steels where the key phase changes occur[152].

Characterisation

A number of dimensionless numbers have been used to characterise temperature:[151]

- **Marangoni Number** - Ratio of surface tension force to viscous force. Higher values represent higher circulation and larger weldpools. Marangoni number,

$$M_a = -\frac{d\gamma}{dT}\frac{L\Delta T}{\mu\alpha} \tag{3.15}$$

where α is the thermal diffusivity, μ is the viscosity ΔT is the difference between solidus temperature and peak pool temperature and L is the characteristic length and can be taken as the depth of the pool.

- **Peclet Number** - Measures sig
higher convective heat transfe

where U is the characteristic l

Other indicators are available in the literature but are not relevant to this work[151].

3.2.4 Stability

A key phenonema within AM is the breakup of the melt pool which typically occurs at higher scanning speeds as well as for larger melt pools[153]. Comparatively minor changes make a significant difference to the form and microstructure of the result[154].

3.2.5 In-Situ Measurement

Similar difficulties to measuring in-situ temperatures, Section 3.2.3, are found in measuring physical processes. While much discussed, the available techniques are accompanied by significant uncertainty[155].

The majority approaches rely on cameras operating in either the visible or infrared regimes[156–158]. Image recognition techniques are used to add distortion and melt pool estimation, feedback automation loop and error detection[159–165].

Measurement of pool topography can be estimated using digital holographic microscopy (CHM), fringe projection and interferometry and photogrammetry[166]. The material composition can be estimated using acoustic spectroscopy (AS)[167, 168]. Ejecta and plume dynamics can be understood using traditional high-speed camera or holographic interferometry (HI)[169–172]. neutron diffraction (ND) has also been used[173, 174].

A large array of post facto measurements are available. Topography can be measured at a range of scales using coherence scanning interferometry (CSI), vertical scanning

interferometry (VSI), phase scanning interferometry (PSI), focus variation microscopy (FVM), x-ray coherence tomography (XCT), atomic force microscopy (AFM) or coordinate measurement machines (CMMs). Material structure can be determined by: confocal microscopy, ramen spectroscopy (RS), x-ray Photoelectron Spectroscopy (XPS), energy dispersive x-Ray spectroscopy EDS[175]

3.3 Defects

There are many causes of defects within additive manufacture.

3.3.1 Porosity and Voids

The porosity of AM parts is often high for a number of reasons[176]. Firstly due to insufficient heat penetration in a region which leaves an area of unmelted powder or a lamination insufficiently bonded to the previous layer[177]. Secondly due to gas trapped in the powder during the atomisation process[178]. Thirdly due to too high a power density causing the weld to operate in *keyhole* mode with the walls periodically collapsing around the weldpool[179]. Fourthly due to the inert shielding gasses getting being trapped in the welded laminate[180].

Total porosity can be determined by buoyancy tests[181]. Cross-sectional porosity can be determined by microscopically examining a section though blurring issues occur[182]. Full 3D analysis can only be achieved using scanning electron microscopes (SEMs), x-ray coherence tomography (CT) or synchrotron radiation micro-tomography (SRμT)[183–185]. These final approaches are highly expensive, however.

Empirical estimations from material response can be made for real-time measurements but no current technique exists for accurate estimation.

3.3.2 Surface Roughness

Surface roughness measurements are a discipline in themselves. The average surface roughness R_a is a commonly used parameter and is given as

$$R_a = \frac{1}{N}\sum_{n=1}^{N}|f_n| \tag{3.17}$$

where values f_n represent N height measurements evenly recorded along a length[186].

Typical R_a requirements of a high quality finished part are in the order of 1µ[187]. As these cannot normally be achieved, finishing methods including grinding, machining, shot peening (SP) and hot isostatic pressing (HIP) are required[107].

The finished roughness is highly dependent on a range of parameters. While some like particle size distribution (PSD) and powder morphology are obvious, others are less so. These include scanning strategy, orientation, location on the bed and travel height[188].

There are three main sources of roughness: layer height, balling and powder. The layer height adds a *stair* effect to sloped surfaces. For layer thickness t_i and slope angle θ this is equal to[112]

$$R_a = 1000t_i \sin\left(\frac{90-\theta}{4}\right)\tan(90-\theta) \tag{3.18}$$

The second major source of roughness is from weld instability. These can lead to *balling* due to the Rayleigh instability caused by surface tension[92]. Thirdly, insufficient heat input can result in loose powder sticking to the surface of the object without being melted[189].

Design of the component as well as position on the bed is currently the most productive method of reducing roughness[190]. An areal approach may lower the energy and processing cost of thinner layer sizes and should allow for a more even heat distribution.

3.3.3 Cracking

Cracking occurs in AM primarily due to three reasons. Firstly, *delamination* occurs when two individual levels fail to bond or the residual stresses exceed the strength of the material. This can be modelled theoretically.

The second case is grain boundary cracking where the weldpool, thermally contracting during cooling, pulls away from the cooler surrounding material. While this can be controlled for experimentally, the effect is too localised for incorporation into existing models[191].

The third reason for cracking is in the partially melted zone (PMZ) During the transition from liquid to solid, a volumetric contraction occurs, causing the grains to separate from the surrounding liquid material[192]. Slower cooling in areal approach means better results.

3.3.4 Residual Stress

The steep temperature gradients between fresh material and the substrate lead to high residual stresses. Not only does this cause cracking, Section 3.3.3, but it causes major distortion and geometric errors. While possibly the most important cause of defects, this is also one of the hardest to predict and computational modelling is expensive. An areal approach should result in much slower cooling times[193].

Computational approaches typically involve solving for the thermal flow and then treating that as a thermal load in a second model[194]. This decoupling of solutions can result in calculation times an order of magnitude lower[195].

An incremental relationship is used for stress/strain to account for the steep temperature gradients involved.

$$d\sigma = D^{EP} \cdot (d\varepsilon^T + d\varepsilon^E + d\varepsilon^P + d\varepsilon^V) - D^E \cdot (d\varepsilon^T + d\varepsilon^V) \tag{3.19}$$

where $d\varepsilon^T$, $d\varepsilon^E$, $d\varepsilon^P$ and $d\varepsilon^V$ represent thermal, elastic, plastic and volumetric strains respectively. The elastic and elasto-plastic stiffness matrices are given by D^P and $D^E P$. The elastic stiffness matrix is determined by Poisson's ratio and Young's modulus for the material while D^{EP} is taken from an experimental knowledge of the material. The thermal strain $d\varepsilon^T$ is taken from the thermal model and the volumetric strain $d\varepsilon^T$ equal to the phase change volume change[196, 197].

The relevant modelling parameters have been extensively covered for titanium,[198] stainless steel,[199] carbon steel[200] and nickel-based alloys[201].

3.3.5 Fatigue

Thermal cycling and repeated deformation are likely to play a significant role in AM part failures. There appears to be little to no quantitative results in the literature on the topic and no models adequately account for fatigue. Further research is required.

3.3.6 Separation and Loss of Alloys

Differing vaporisation temperatures and volatilities of alloying elements mean that varying percentages of alloying materials can be lost. This change in composition results in microstructure changes. An areal approach to heating would allow for greater control of temperature and reduced vaporisation. Composition of the materials before and after AM can be determined with electron probe microanalysis (EMPA) and energy dispersive spectroscopy (EDS) and alloy compositions compensated for[202]. The literature suggests that around a 10-15% change in composition of secondary elements such as magnesium and aluminium is common[203]. As nucleation is a fundamental solidification process, predictive modelling is challenging[204].

The Langmuir equation can be used to estimate the vapor flux J_n

$$J_n = \frac{\lambda_c P_n}{\sqrt{2\pi M_n T}} \tag{3.20}$$

where the molecular weight and equilibrium vapour pressure for element n are given by M_n and P_n respectively. The value of λ_c is usually determined experimentally as theoretical models are only valid for operation in a vacuum. The main options for reducing λ_c is in reducing the weldpool diameter and peak temperatures[205, 206]. It is expected that a lower temperature areal process will greatly reduce the alloy vaporisation.

3.4 Conclusion

This chapter has introduced some of the complexities of modelling additive manufacture processes and the technique used. Additionally, it has highlighted some of the particular challenges faced by L-PBF systems and how they can be handled. In later chapters we will return to this topic when with a discussion of the design of a powder system.

Chapter 4

HoloGen

Chapter 4

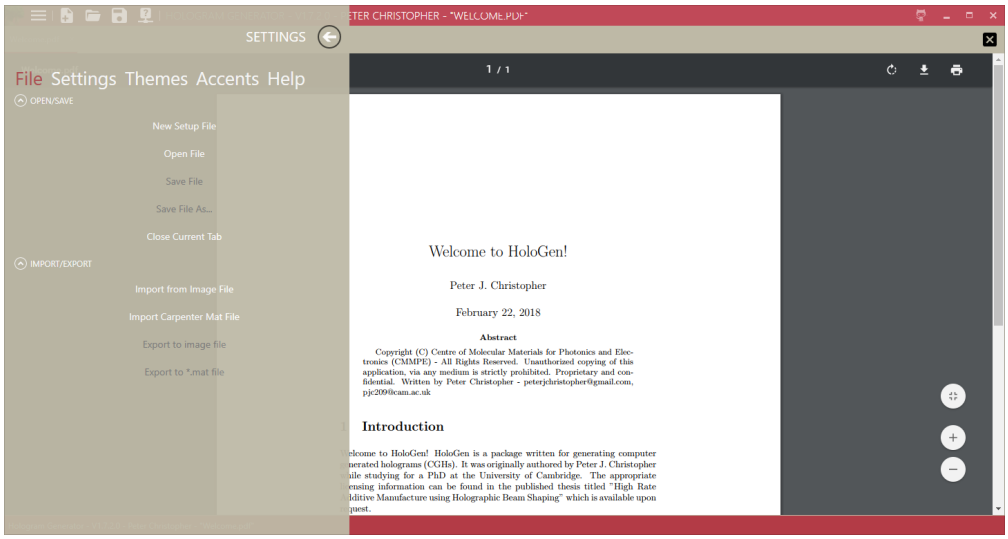
HoloGen

4.1 Introduction

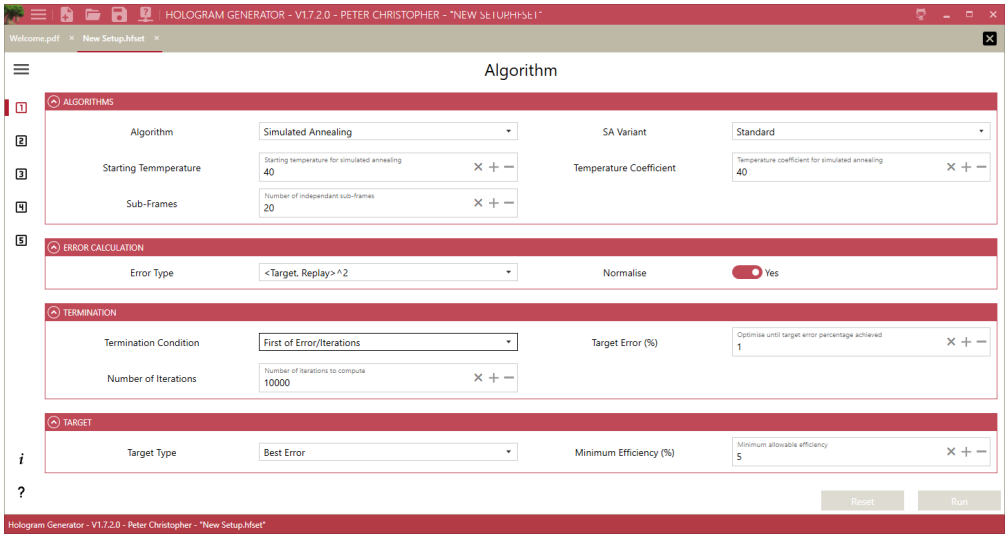
In order to develop the novel algorithms discussed in Chapter 4.3, an application suite was developed for benchmarking. Known as HoloGen, the suite is built on top of a custom parameter framework in C# and WPF. The algorithms are implemented in Cuda[207]. The application at the time of writing ran to $76,000$ lines of code over 38 new libraries. Version 1.15.6 of HoloGen also imports 25 external libraries.

The HoloGen application is built on a MVVMA architecture. This is a standard Model-View-ViewModel (MVVM) framework common in C# windows presentation framework (WPF) applications with an additional algorithms level written in a more traditional procedural or functional style on top of an Nvidia Cuda architecture interfaced in C++[207]. The general application levels are shown in Figure 4.7. This shows the three application levels: the user interface level; application level and algorithm level. These all depend only on levels beneath them and have their own independent imported libraries. A more detailed breakdown is shown in Figures 4.8, 4.9 and 4.10.

The cuFFT library from NVidia is used to perform the FFT element of the algorithms due to its high performance[208]. This is built on top of a Cuda framework with Thrust wrapper. Previous researchers have also used the FFTW library[209, 210], OpenCV[211], OpenMP[212], the computational wave optics (CWO)[213–215] and Intel Math Kernel

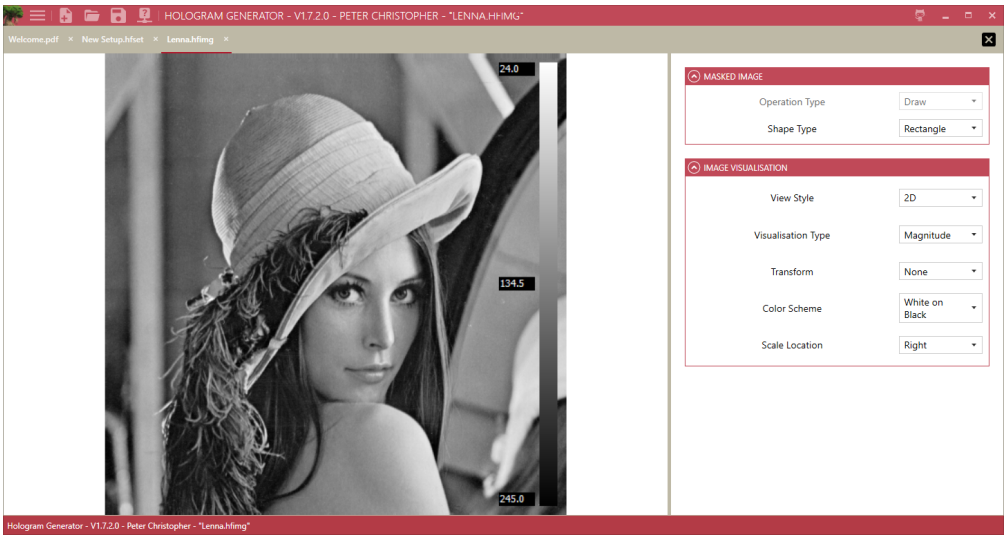


(a) Welcome screen with integrated browser

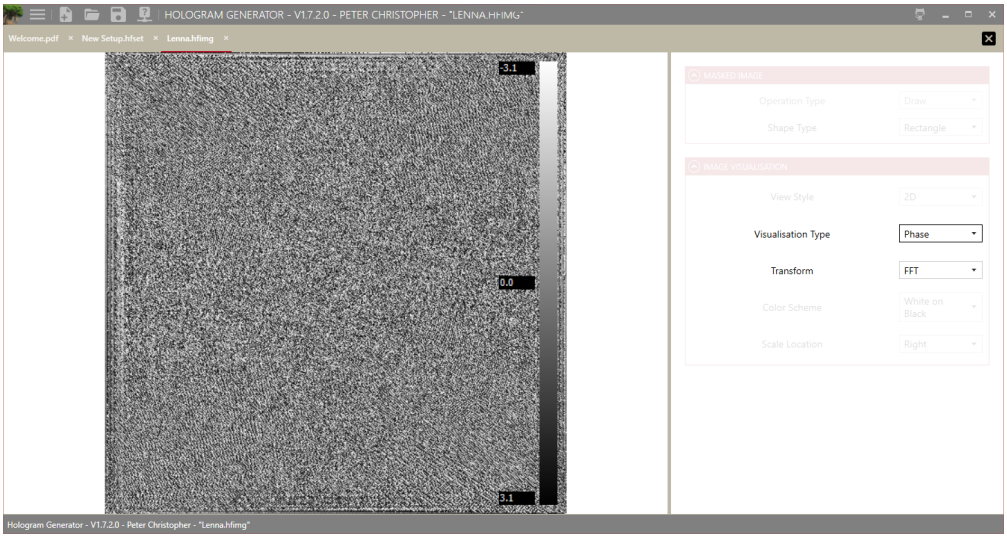


(b) Algorithm options

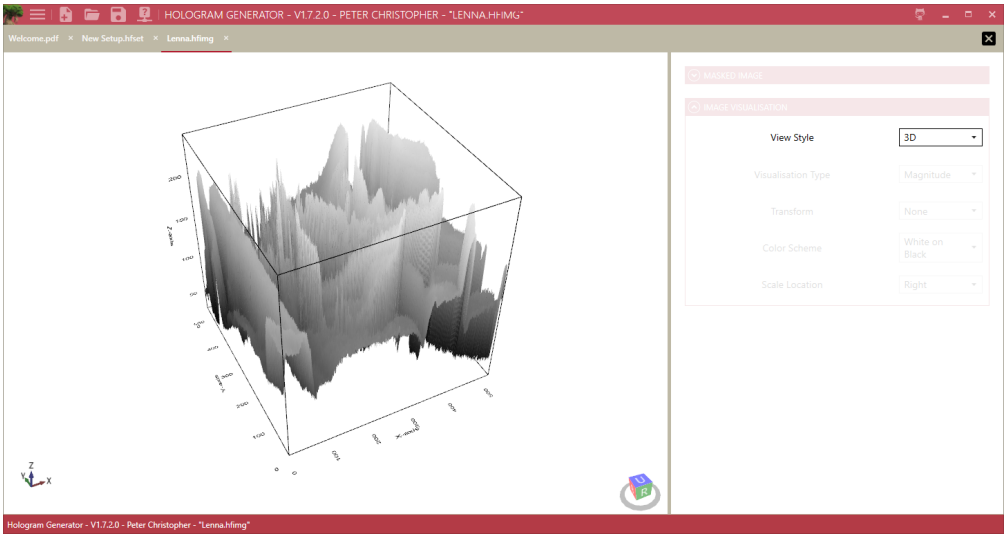
(c) Algorithm options with Chinese translation



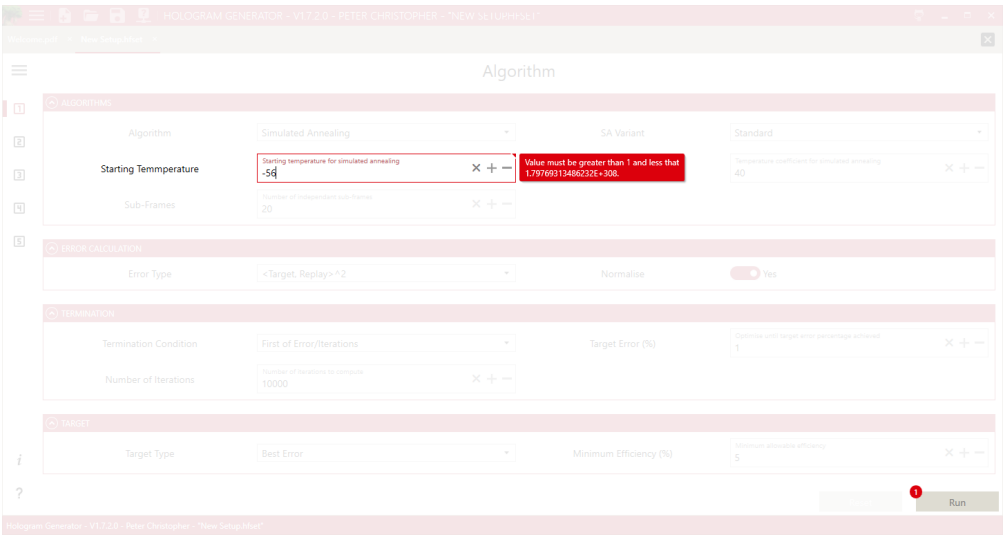
(d) Image visualisation



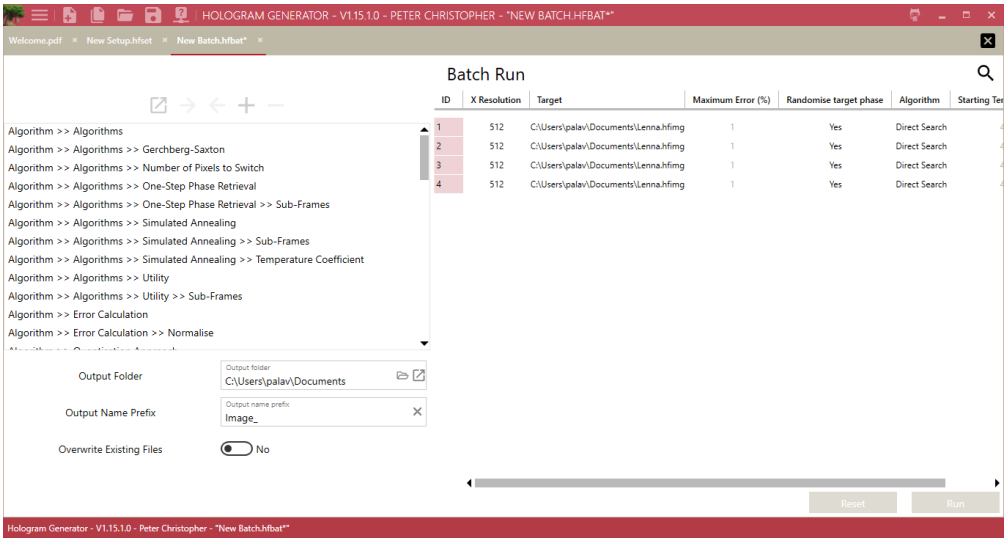
(e) Hologram visualisation



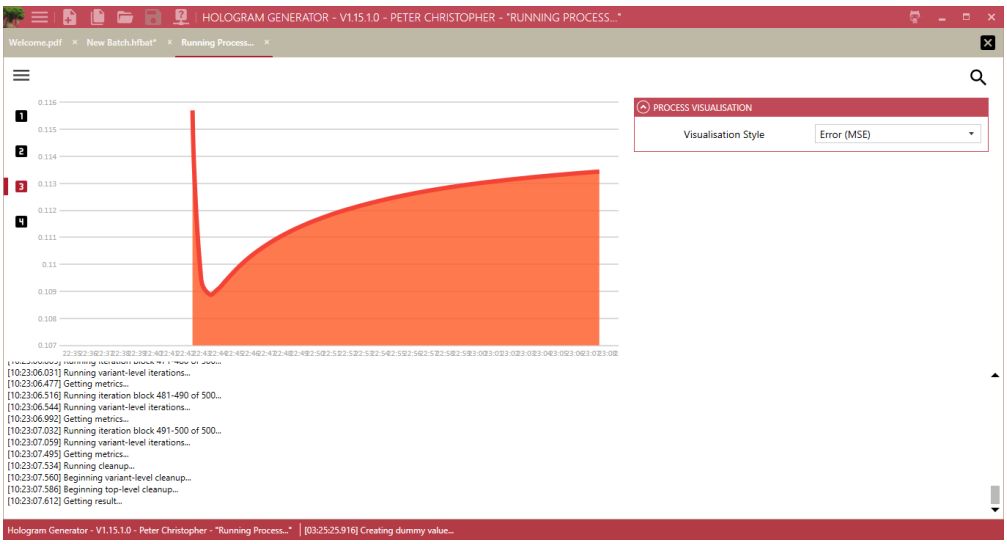
(f) 3D visualisation



(g) Error reporting



(h) Batch processing



(i) Real-time algorithm reporting

Fig. 4.1 HoloGen screenshots.

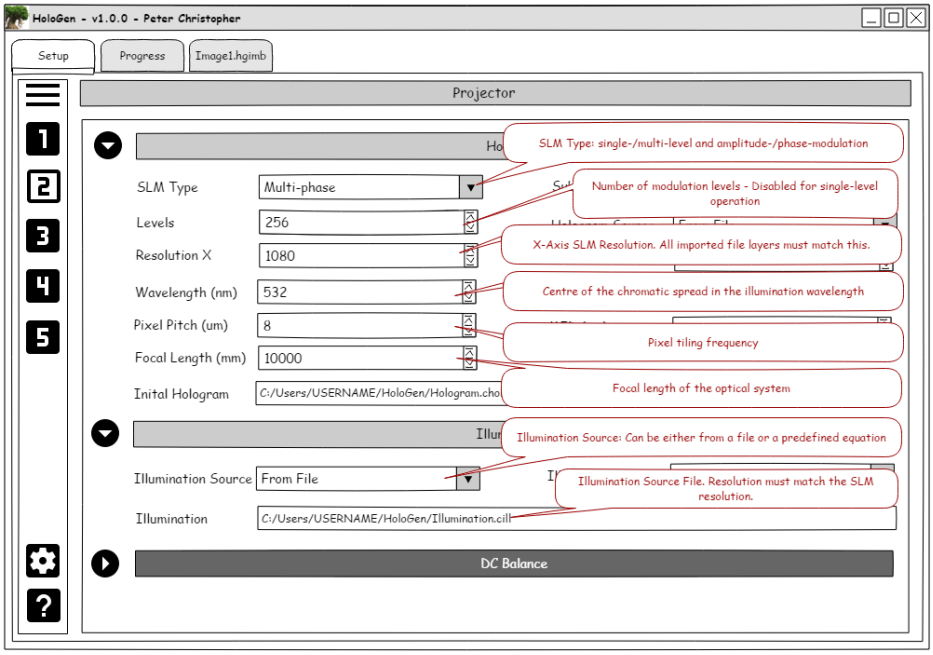


Fig. 4.2 Part of the initial HoloGen specification.

(MKL)[216] libraries as well as custom implementations in Matlab[217–222], C/C++[223, 212, 216, 209, 210, 213–215] and Python.

HoloGen depends on a number of third-party libraries[224–241]. Most are licensed under the MIT license[242–250, 236, 251] with several licensed under the Lesser General Public License[252–255], several under the Apache License 2.0[256–258, 240] and two under the three-clause BSD license[259–261]. . These are all highly permissive licenses. Xamarin.Forms, C# and WPF come with appropriate usage licenses as part of Cambridge University Visual Studio package while Cuda comes with an appropriate license and EULA[262, 263].

HoloGen includes a number of novel elements. A reflection based parameter framework allows for persistent parameter models with limited code reuse. A dynamic module system interfaces with Cuda implemented algorithms allowing for real-time Cuda compilation on host machines improving performance and utilising newer features of newer graphics cards. 3-dimensional visualisation techniques can be used while viewing the generated hologram statistics. Fourier transform functionality is incorporated directly into the viewer. A tabular batch processing framework allows for multiple operations to be scheduled for a background run. Advanced tabulation allows for comparison of different holograms. New image and file types are introduced to handle the additional information available and all results are tagged with parameter metadata used to ensure traceability. Provisional translations into

French, Spanish, German and Chinese are available as well as complete help documentation. An included Chromium browser allows for integrated reporting features. HoloGen also introduces a new file format for image representation due to the limitations of existing formats[264].

Part of the initial specification for the application is shown in Figure 4.2. Screenshots of the application are shown in Figure 4.1. Figure 4.1a shows the main introduction screen and a selection of the setup options are shown in Figure 4.1b with example translation in Figure 4.1c. Figures 4.1d, 4.1e and 4.1f show the image visualisation features. Figure 4.1g shows the error reporting processes while the batch processing framework is shown in Figure 4.1h. Finally, Figure 4.1i shows the real-time progress reporting.

HoloGen also includes a full benchmarking suite for measuring and graphing the performance of different holographic algorithms. An example of benchmarking the Gerchberg-Saxton algorithm from Section 2.4.3 is given in Appendix B.

For more information on the application architecture consult Appendix G.

4.2 Selected Details

While HoloGen application, Figure 4.1, has targeted traditional structure for ease of extension, a number of structural and implementation features deserve mention. Additional detail is packaged with the source code.

4.2.1 Graphical User Interface

The graphical user interface (GUI) is based on the windows presentation framework (WPF). WPF in turn, uses the extensible application markup language (XAML) to define the user interface components. Like the majority of modern GUI packages, WPF encourages *binding* where elements in the *view* layer are bound to properties and collections in the *viewmodel* layer. This approach allows for two way data flow and removes much of the filler code found in older primarily event driven architectures such as WinForms. This approach also allows for easy runtime injection and extension, meaning that GUI portions are only loaded when required.

4.2.2 Reflection Parameter Hierarchy

HoloGen uses a custom reflection based parameter and command system. This is in contrast to the XML parameter sheet systems in widespread use. Instead of the parameter types and interactions being defined in parameter sheets which are parsed at runtime, the parameter system is coded into the C# directly. This significantly reduces the runtime overhead as well as improves the error checking available at compile time. The downside is an increased architecture exposure of the parameter hierarchy.

4.2.3 Interop

For fast and easy transfer of large images to the C++ subsystem, a three level architecture is used. The use of managed C++ increase the structural complexity but allows the C# application layer to be ignorant of the dynamic link library (DLL) interface. The use of the native C++ layer allows the use of Nvidia Thrust tools as class members.

4.2.4 Fast Fourier Transforms

The majority processing factor in any holographic system is the two-dimensional Fourier transform. Tests found that the FFT calculation or update step took over 98% of the runtime for all algorithms on system input and output operations were excluded. As a result, any implementation is heavily dependent on the FFT library used.

HoloGen currently uses cuFFT, Nvidia's implementation for their graphical processing units (GPUs), due to its high reliability and performance[208]. A graph of the performance of cuFFT against resolution is shown in Figure 4.3 along with the idealised $O(N^2 log(N)^2)$ trend line.

4.2.5 Floating Points

The IEEE standards define 32-bit and 64-bit floating point numbers, represented in C++ by `single` and `double` values. Less widely used is the 16-bit floating point[265]. 16-bit numbers are ideal for GPU based computation and in particular holography. Real-world

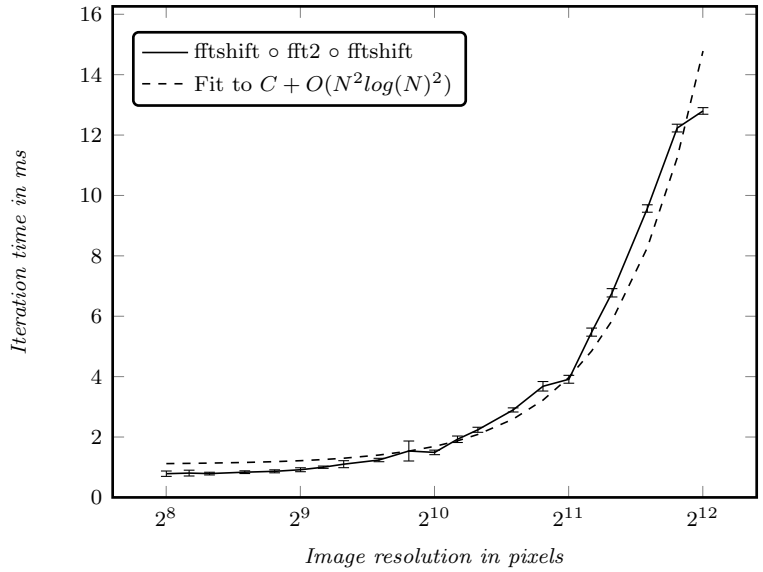


Fig. 4.3 Performance of cuFFT for differing square image resolutions. Error bars show the 2σ confidence interval measured from 100 independent runs of 1000 pairs of FFTs and IFFTs

image formats are typically 8-bit per colour meaning that a 16-bit floating point, when scaled correctly, can more than accommodate the necessary information while significantly improving performance. The scaling element is the key for 16-bit operations were care must be taken to normalise all FFT operations to reduce unexpected errors and overflows.

HoloGen is capable of being compiled in 16-, 32- and 64- bit versions with the application performance being approximately proportional to the reciprocal of the number of data bits. HoloGen also automatically scales every image in order to increase accuracy at low bit levels.

4.2.6 Templatisation

The standard version of HoloGen tracks properties such as the illumination fields that are not necessary in some applications. By making significant use of the C++ template syntax, this can be tuned at compile time. This allows compile time flexibility in the required algorithm portions while still offering runtime performance.

This is combined with the Nvidia runtime compilation (NVRTC) which allows users the ability to modify algorithms at runtime. This is not currently exposed in the GUI for HoloGen but is available in the application programming interface (API).

```
template<bool FullCircle, typename FloatType, typename IntType>
struct quantiseDiscretePhase {
    private:
    constexpr float _pi = 3.14159265359;
    inline constexpr const FloatType ConstrainSLM(
        const FloatType& diffArg, constexpr FloatType illumArg, constexpr FloatType illumAbs) {
        if constexpr (!FullCircle) {
            while ((diffArg - _maxSLMArg) > _pi*2)
                diffArg -= _pi*2;
            while ((diffArg - _minSLMArg) < -_pi*2)
                diffArg += _pi*2;
            if (diffArg > _maxSLMArg)
                return thrust::polar<FloatType>(
                    illumAbs,
                    illumArg + (diffArg < _wrapMaxSLMArg?_maxSLMArg:_minSLMArg));
            if (diffArg<_minSLMArg)
                return thrust::polar<FloatType>(
                    illumAbs,
                    illumArg + (diffArg > _wrapMinSLMArg?_minSLMArg:_maxSLMArg));
        }
        return diffArg;
    }
    const FloatType _minSLMArg;
    const FloatType _maxSLMArg;
    const IntType _levels;
    const FloatType _spac;
    const FloatType _wrapMinSLMArg;
    const FloatType _wrapMaxSLMArg;
    public:
    quantiseDiscretePhase(
        const FloatType minSLMArg, const FloatType maxSLMArg, const IntType levels) :
        _minSLMArg(minSLMArg),
        _maxSLMArg(maxSLMArg),
        _levels(levels),
        _spac((_maxSLMArg - _minSLMArg)/(_levels - 1)),
        _wrapMinSLMArg(maxSLMArg + fmod((maxSLMArg - minSLMArg), _pi*2) / 2.0),
        _wrapMaxSLMArg(minSLMArg - fmod((maxSLMArg - minSLMArg), _pi*2) / 2.0) {
    };
    __device__ constexpr const thrust::complex<FloatType> operator()(
        const thrust::complex<FloatType>& input) {
        const auto inputArg = Globals::Arg(input.real(), input.imag());
        const auto diffArg = ConstrainSLM(fmod(inputArg, _pi*2), 0, 1);
        const auto discArg = _minSLMArg + _spac * roundf((diffArg - _minSLMArg)/_spac);
        return thrust::polar<FloatType>(1, discArg);
    }
};
```

Fig. 4.4 HoloGen quantisation operator for nearest-neighbour quantisation for a discrete phase level SLM.

4.2.7 Example

The code listing in Figure 4.4 demonstrates a number of these principles in action. The `struct` shown, `quantiseDiscretePhase`, exposes the `()` operator. The Thrust library is used to call this as shown in Figure 4.5 where Thrust handles the memory management of calling the `quantiseDiscretePhase` operator on its arguments. Properties such as `FloatType` and `IntType` allow for changing the numerical representation at runtime while use of the `if constexpr` syntax from C++17 allows for unwanted execution pathways to by ignored. Modern C++ allows for significant flexibility between runtime (`const`) and compile time (`constexpr`) constness. By changing the `DerivedPolicy`, it is possible to compile the application for CPU or for GPU operation.

```
thrust::transform(
    DerivedPolicy,
    input.begin(),
    input.end(),
    thrust::make_zip_iterator(
        thrust::make_tuple(IlluminationPhases->begin(),
        IlluminationMagnitudes->begin())),
    input.begin(),
    quantiseDiscretePhase<FullCircle,FloatType,IntType>(MinSLMValue, MaxSLMValue, Levels));
```

Fig. 4.5 Calling an operator using Thrust.

4.3 Validation

Figure 4.6 shows a binary-phase OSPR image generated using HoloGen for a 512×512 pixel image showing the generated hologram (left), target image (top right) and result (bottom right). The poor reproduction quality is due to the limitations of the optical setup used.

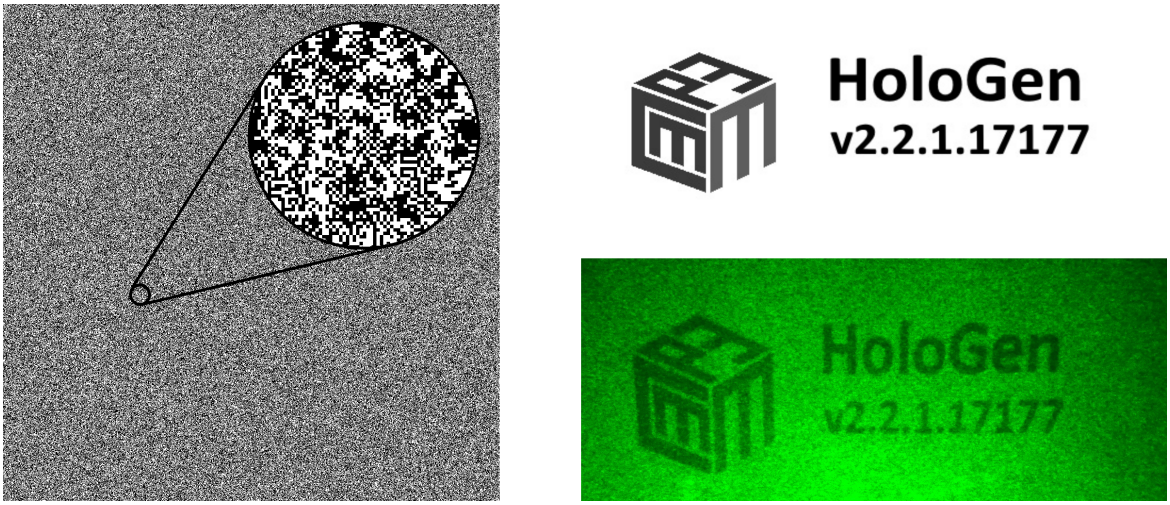


Fig. 4.6 Binary phase hologram generated with HoloGen including target image (top right), single binary subframe (left) and measured result (bottom right). Captured using a Canon 5D Mark III with a 24-105mm lens.

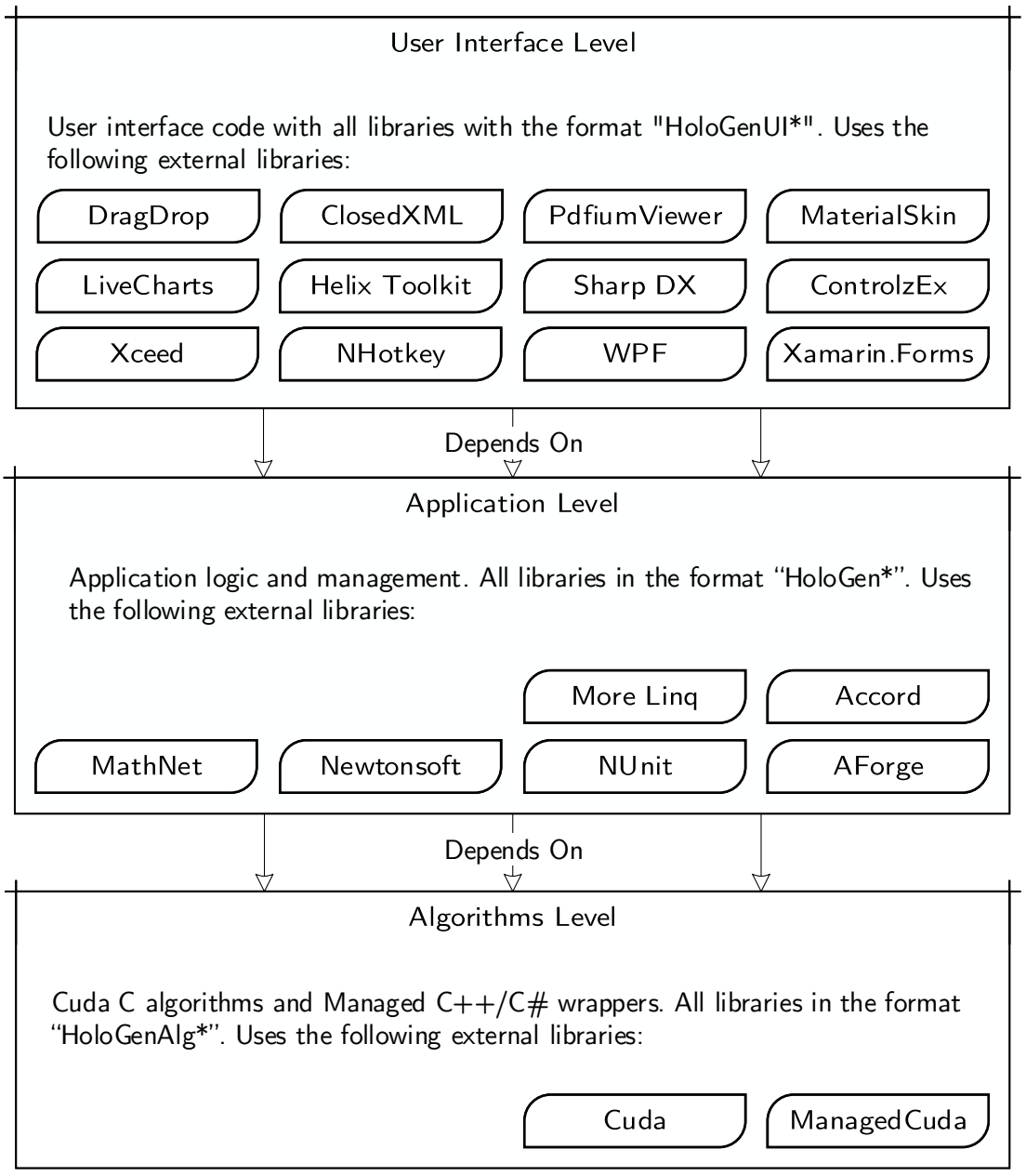


Fig. 4.7 HoloGen application levels. c.f. Figures 4.8, 4.9 and 4.10.

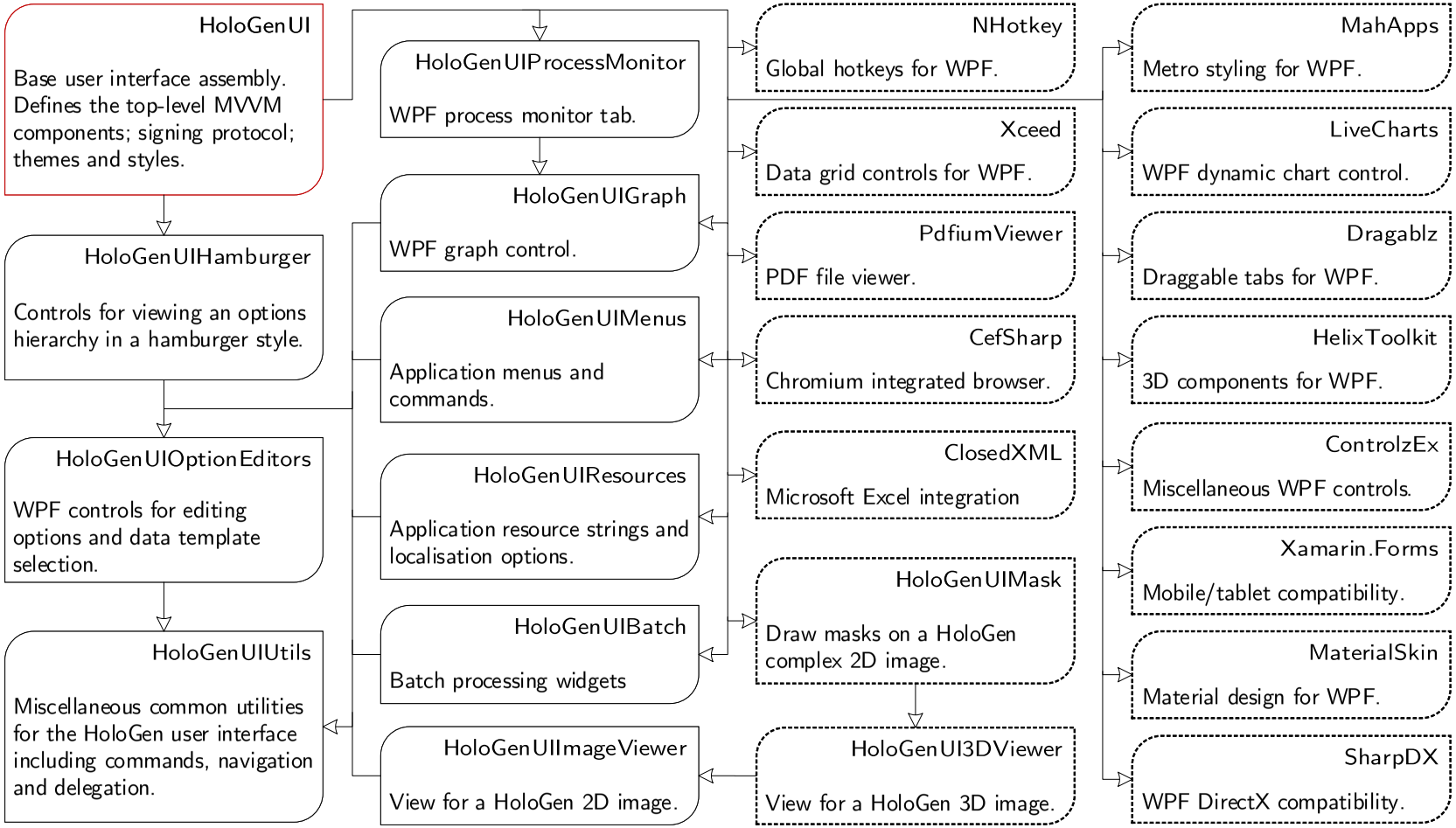


Fig. 4.8 HoloGen user interface level. c.f. Figure 4.7.

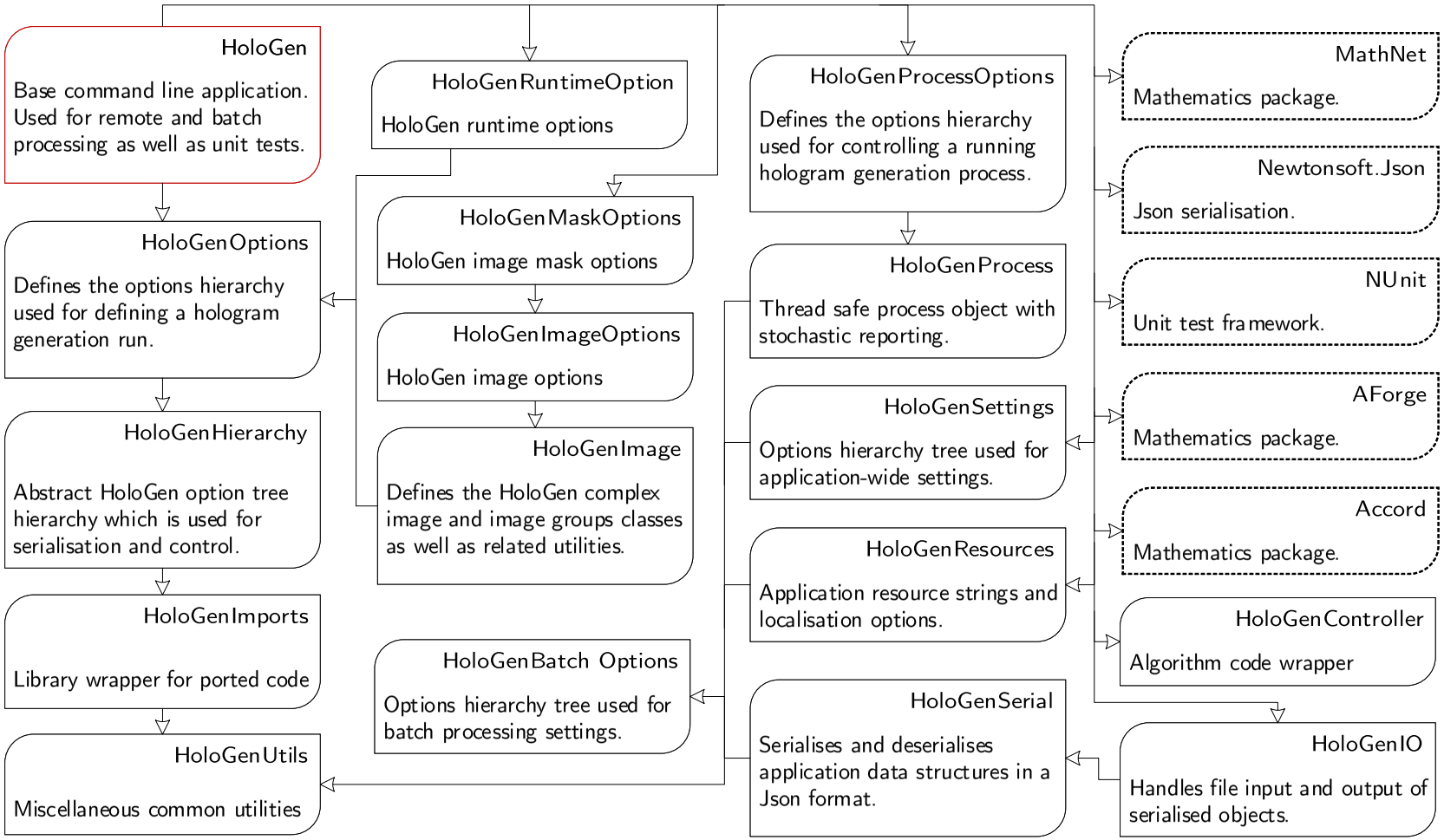


Fig. 4.9 HoloGen application level. c.f. Figure 4.7.

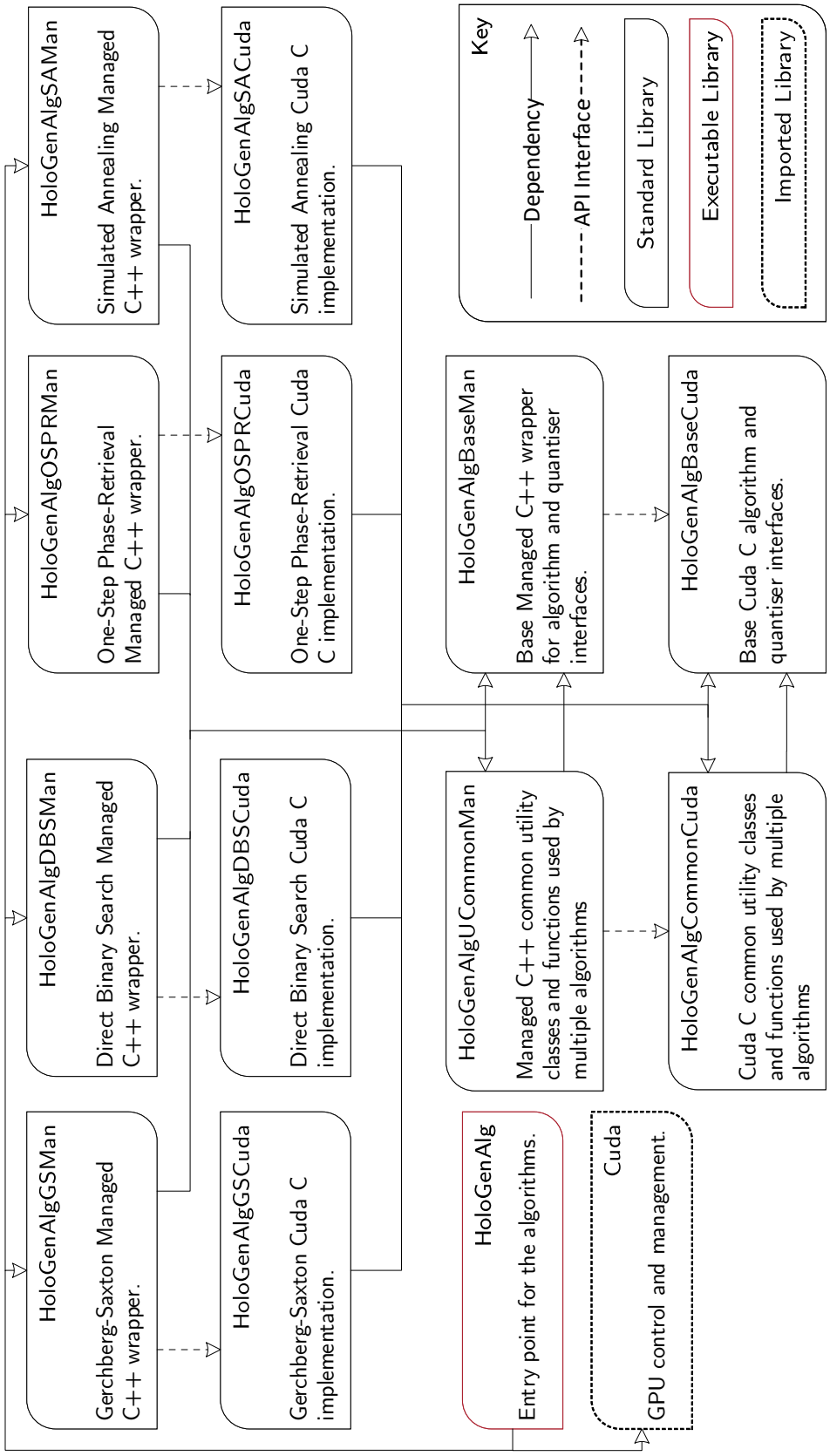


Fig. 4.10 HoloGen algorithm level. c.f. Figure 4.7.

Chapter 5

Improved Hologram Generation Algorithms

Chapter 5

Improved Hologram Generation Algorithms

5.1 Introduction

As part of this work, a number of novel algorithms and techniques for hologram generation were developed. This section introduces the experimental system used before the individual algorithms are presented in Sections 5.2 to 5.6.

5.1.1 Systems

In order to develop these new algorithms a Jasper JD8714, Table 6.1, was used along with a simple polariser/analyser combination as shown in Figures 5.2 and 5.1. This arrangement allows for the SLM to be used as either a binary or multi-level device as well as either in phase or amplitude by using the correct configurations of polarisers. The SLM is illuminated using a 532nm laser diode spatially filtered using a single mode fibre.

The primary aim of the system was to be as simple and as flexible as possible so as not to obscure the effect of the algorithms on the final result.

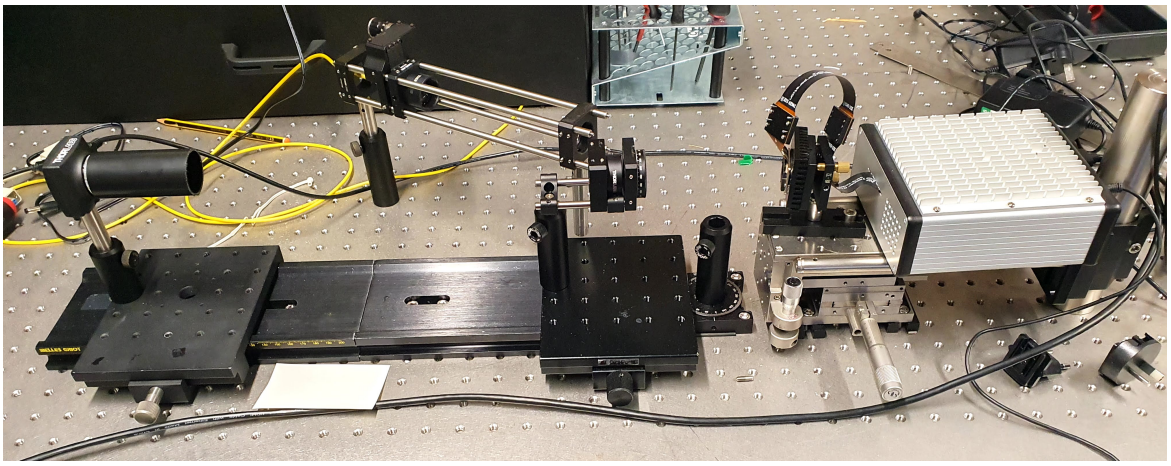


Fig. 5.1 Algorithm experimental confirmation photograph

5.1.2 Algorithms

Five new algorithms and algorithm families are presented in the following sections with the following use cases:

- **Sorted Pixel Selection** - Section 5.2 - Primarily useful as an optimisation on search algorithms including direct search and simulated annealing [266].
- **Single-Transform Time-Multiplexed** - Section 5.3 - An improvement on older algorithms such as one-step phase-retrieval for realtim operation with high frame-rate binary devices [267].
- **Holographic Predictive Search** - Section 5.4 - A family of algorithms that offers order of magnitude performance improvements over search algorithms at the expense of a more complex mathematical formulation [268, 269].
- **Sympathetic Quantisation** - Section 5.2 - Useful in more modern real-time generation such as headsets where multi-level devices are used [270].
- **Linear Time** - Section 5.6 - Targeted at phase sensitive replay fields, this offers a linear time algorithm for generation of holograms [271].

In addition, Appendices A-D offer further discussion of algorithmic elements.

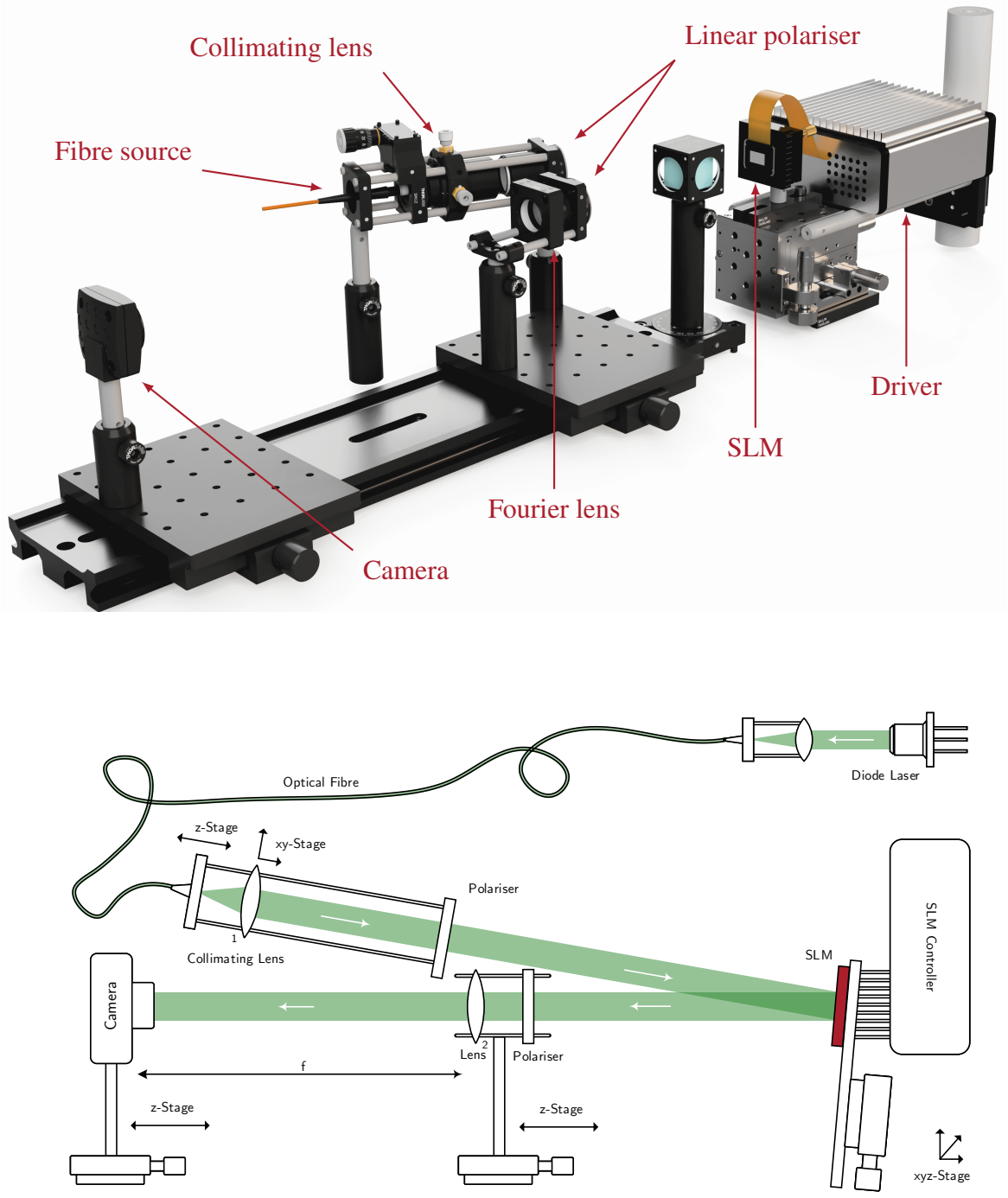


Fig. 5.2 Algorithm experimental confirmation configuration render (above) and schematic (below)

5.2 Sorted Pixel Selection[1]

5.2.1 Introduction

This section, in particular, discusses *sorted pixel selection*, an improvement to the existing direct search and simulated annealing algorithms introduced in Sections 2.4.4 and 11.

5.2.2 Quantisation and Initial Guess

Search algorithm performance depends heavily on the initial guess for the diffraction field hologram. By far the dominant approach is to back-project the target using an inverse DFT. The limited modulation abilities of SLMs, Section 1.3.2, means that the initial back projection requires *modulation* and *quantisation* steps in order to adapt the complex valued function to the constraints of the display device. Each modification of a pixel value has an associated change in result, usually causing an increase in MSE.

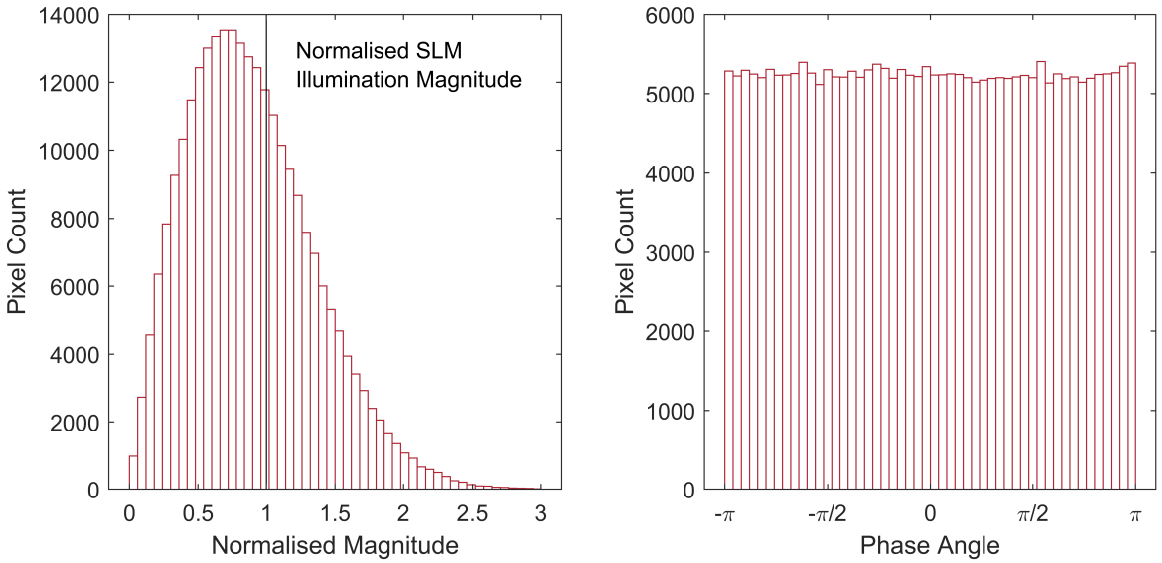


Fig. 5.3 Spread of magnitudes (left) and angles (right) of a back projected version of *Mandrill*

[1]The contents of this section were previously included as part of the following publication:

Peter J. Christopher, Jamie Lake, Daoming Dong, Hannah Joyce, and Timothy D. Wilkinson, "Improving holographic search algorithms using sorted pixel selection", *Journal of the Optical Society of America A* (2019), Volume 36, Issue, pp. 1456-1462, DOI: 10.1364/JOSAA.36.001456

5.2.3 Heuristic Observations

The back projection of a phase-randomised target image shows a spread in pixel magnitudes and phases, Figure 5.3. For the case of the standard *Mandrill* test image, Figure 5.4 (left), the normalised values are shown in Figure 5.4 (right) where normalisation is taken as the equivalent of an illumination field of unit magnitude. The quantisation step will constrain these values to the SLM. In the case of phase holography, this will leave phase unchanged but will set all magnitudes equal to unity while in amplitude holography the amplitudes are set to the real values of the combined complex numbers.

The artificially induced symmetry in the test image is to account for the rotational symmetry inherent in binary devices. While this is not ideal in terms of natural images, it is sufficient for the purposes of demonstration.

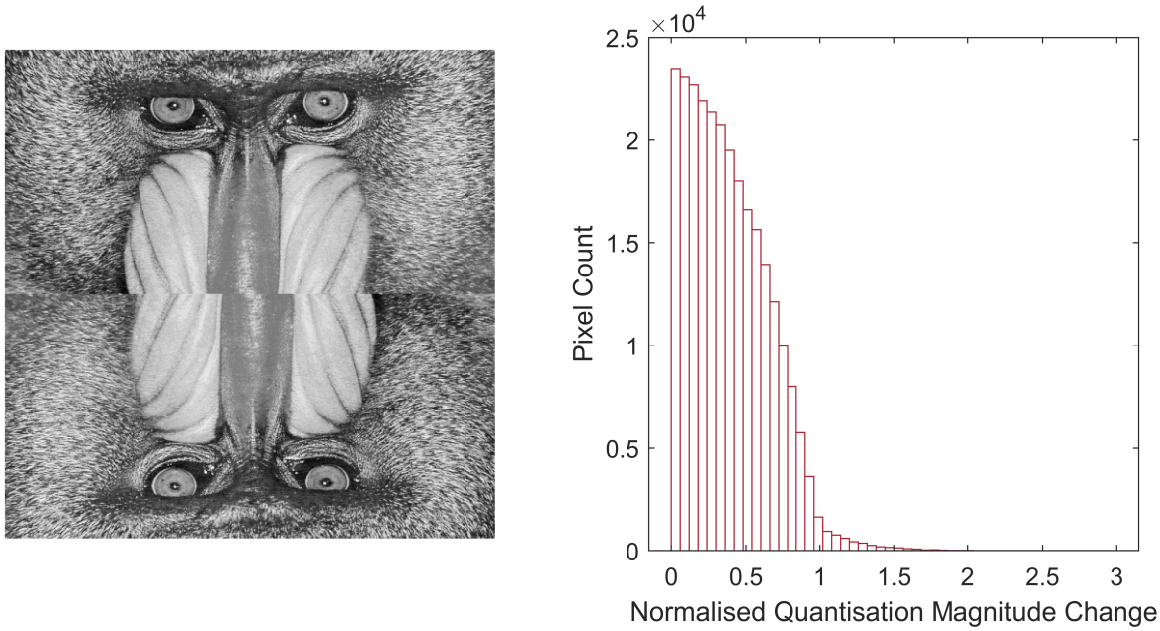


Fig. 5.4 Test image *Mandrill* with artificially induced symmetry at 512 × 512 resolution (left) and spread of pixel magnitude changes due to quantisation for display on a pure phase device (right)

The premise of this section is based on the observation that the total change in error of the replay field caused by changing a single pixel is highly correlated to the magnitude of the SLM pixel value change during quantisation. Figure 5.5 shows a scatter plot of the correspondence between independently quantising each of the 2.6×10^5 pixels of a continuous phase hologram and the resultant MSE.

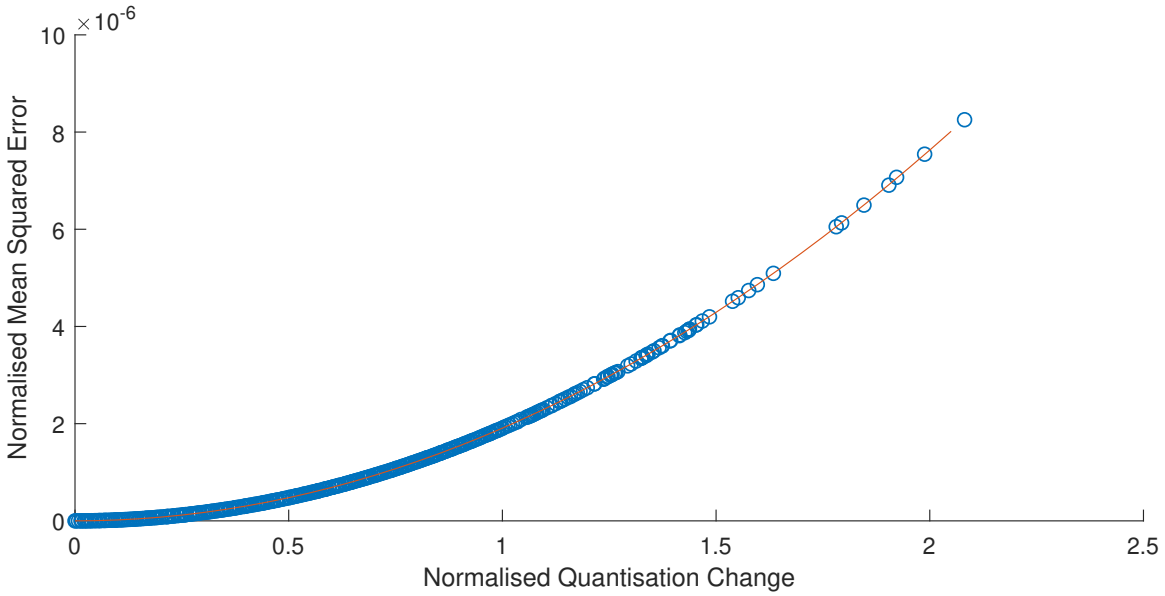


Fig. 5.5 Scatter plot of pixel value changes during quantisation of a continuous phase hologram against resultant error as well as expected square law relationship.

Figure 5.5 also shows the expected square law relationship due to the MSE error metric used. A correlation of $\gg 0.99$ was observed between the trend line and the simulated dataset.

Traditional search algorithms randomise the test pixel selection process. The relationship between change in pixel value during quantisation and the effect on the replay field suggests, however, that convergence will be improved by testing pixels with the greatest change during quantisation as they are heuristically likely to have greatest impact on error reduction.

This approach is here termed sorted pixel selection (SPS) where, instead of randomly selecting test pixels, test pixels are chosen sequentially from a list in order of decreasing quantisation change.

5.2.4 Results

For example, modifying the direct search algorithm in Figure 2.14e to successively test pixels in order of decreasing quantisation change provides the convergence graph shown in Figure 5.6 for a modified algorithm as shown in Figure 5.8. This uses the 512 pixel square *Mandrill* test image on a binary phase SLM and exhibits a 16.5% improvement in error reduction over $200,000$ iterations. In this implementation this is equivalent to a

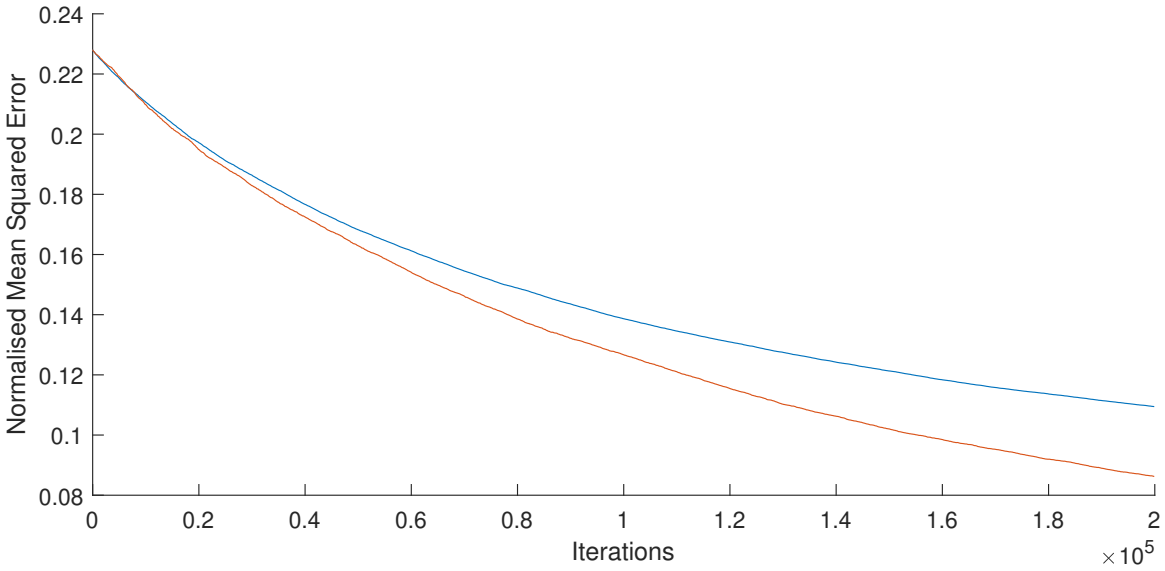


Fig. 5.6 Convergence of direct binary search with random pixel selection (blue) against sorted pixel selection (orange) for a 512×512 pixel *Mandrill* test image being displayed on a binary phase spatial light modulator

900 millisecond runtime as opposed to approximately 10 minutes. In Figure 5.7 shows a comparison of the effect this has on a rotationally symmetric text image.

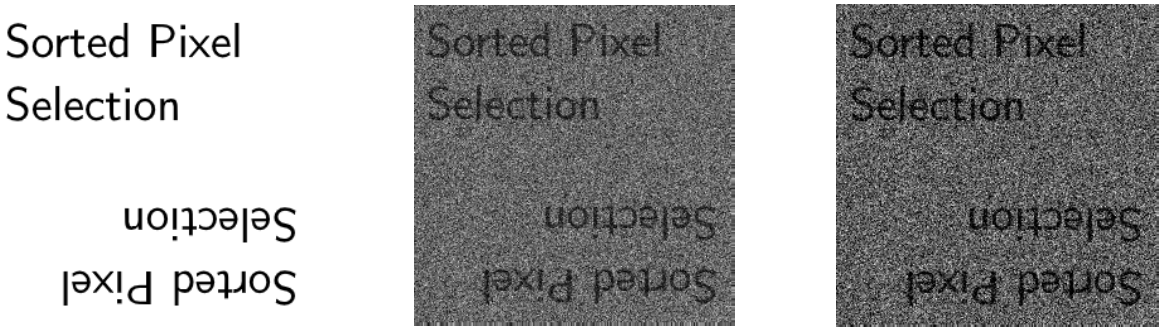


Fig. 5.7 Binary image generated with 10,000 iterations of direct search (centre) and direct search with sorted pixel selection (right) for a rotationally symmetric 256×256 pixel target image (left). There is a 19.1% reduction in the mean squared error, when comparing the righthand image and the lefthand image.

This improvement is not massive but is significant due to the very low cost of implementation. There are many sorting algorithms that operate in $O(n \log n)$ time where $n = N_x N_y$, which is the same complexity as the 2D FFT itself[272]. Combined with their widespread availability makes adding a sorting step at the start of a search algorithm trivial.

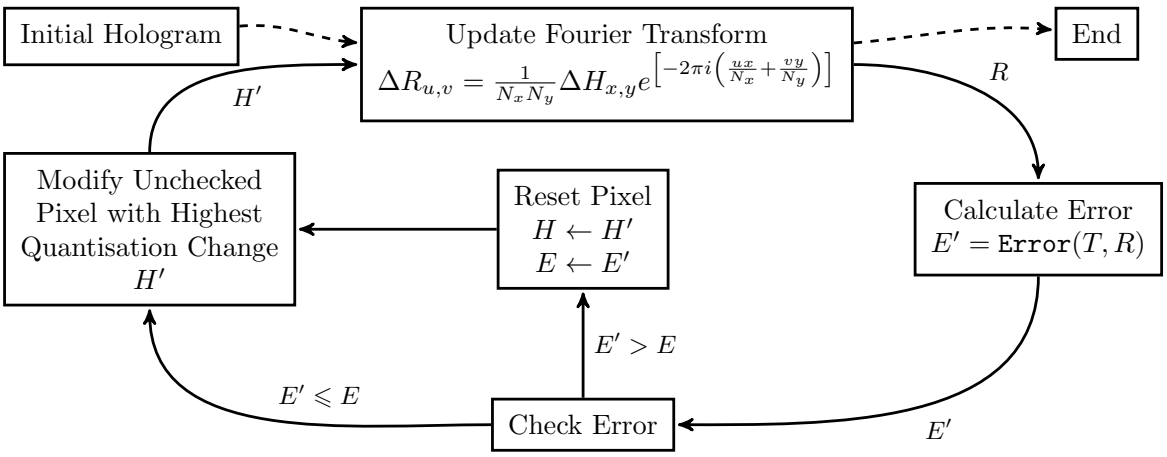


Fig. 5.8 Fast Direct Search with Sorted Pixel Selection

Additionally, it is notable due to its stability. While Figure 5.6 shows the case of showing *Mandrill* on a binary phase SLM using $200,000$ iterations of a direct search algorithm, over 100 tests showed improvements in the range of $[14.7\%, 19.2\%]$ for a range of parameter combinations independent of:

- **Algorithm** - Simulated Annealing showed indistinguishable performance to the Direct Search case.
- **Resolution** - Both algorithms experienced similar relative performance increases for image resolutions between 64 pixels square up to 2048 pixels square.
- **Image** - While *Mandrill* was the primary test image used, tests with other images from the USC-SIPI image database showed similar results[49]. This included tests with the USAF target with a very different initial power spectral density.
- **Iterations** - The relative performance gain was consistent at $\sim 15\%$ regardless of the number of iteration count.
- **Modulation Scheme** - While the graph shown in Figure 5.6 is for the case of a pure phase modulator, no difference in relative performance improvement was observed for the case of a pure amplitude modulator. The case of hybrid phase-amplitude modulation schemes has not yet been rigorously tested.
- **Modulation Levels** - Search algorithms are primarily used in the case of binary holograms as competing algorithms such as Gerchberg-Saxton offer better performance in the multi-level or continuous cases. Running tests for 8, 256 and continuous

Fig. 5.9 Experimental result of taking 10,000 iterations of direct search (left) and direct search with sorted pixel selection (right) for the test image shown in Figure 5.4. Test image used is taken from the USC-SIPI database[49] and is shown in Figure 2.12. The SPS image exhibits a slight reduction in noise when compared to the DS generated image and improved contrast and visual quality.

modulation levels showed no significant change in relative convergence improvement over the binary case.

The technique discussed here has been applied to far-field or Fraunhofer holograms. It is expected to be equally applicable to mid-field or Fresnel holograms as the addition of a phase rotation term does not change the heuristic arguments made earlier.

One final observation of note is that while error reduction showed a $\sim 15\%$ improvement, the total number of pixel changes accepted remained largely the same with the impact of successful pixel modifications rising in the SPS case.

5.2.5 Experimental

Figure 5.9 shows the experimental result of taking 10,000 iterations of DS (left) and SPS (right) for the same target image using the experimental system described in Section 5.1.1 with only 512×512 pixels of the Jasper SLM being enabled to ensure a fair comparison. While this shows a similar change in visual quality between the individual frames with the SPS image exhibiting a slight reduction in noise and improved contrast and visual quality, it must be noted that the dominant sources of image degradation appear are not primarily due to the hologram displayed.

5.2.6 Comparison with Previous Work

This section presented a novel method of using the quantisation change to improve the family of search algorithms used in CGH.

Quantisation has been a focus of improvements in holography since the early 70's and it has been recognised that quantisation changes cause a noise term in the reconstruction for several decades[273].

A number of papers have examined the expected proportion of error that will be caused by quantisation as opposed to other limiting factors such as energy conservation and a number of non-iterative algorithms have been developed for reducing this quantisation noise[274]. Non-iterative approaches have included adjusting point source locations spatially in the diffraction field[275] and the use of dummy areas with variable regions of interest[276].

Common iterative algorithms such as GS deal with quantisation error as part of the expected whole[277]. Some sources have discussed quantisation error independently of other sources. For example, Yang et. al. discussed quantisation error in off-axis configurations[278]. HSAs are typically applied in applications with binary or low quantisation levels and quantisation is known to be a major error source[279]. A number of alternative quantisation techniques such as error diffusion have been used to attempt to reduce this as well as replay field manipulation approaches[278, 280].

No previous work has directly linked the quantisation change with expected error on a pixel-by-pixel basis. Figure 5.5 allows the suggestion of a modification to popular search algorithms that improves performance with extremely low additional overhead.

5.2.7 Limitations and Future Investigation

This section has considered only a small subset of holographic search algorithms. While the proposed approach is postulated to be equally efficacious in other cases, this has yet to be tested. This technique has also only been applied to phase insensitive, binary phase Fraunhofer holograms. Further study is required to examine its applicability to phase sensitive applications, alternative devices and the Fresnel region.

While this section has sorted purely by quantisation change, it is likely that other variables will also have significant impact. It is expected that other features such as position in the replay field are likely to make a significant impact.

Also necessary of investigation is the behaviour when the number of iterations is higher than the number of pixels. Both of these issues can be addressed by use of the development of a probabilistic function that weights a number of parameters including quantisation change during the pixel selection step. Instead of iterating through a sorted list, the function would probabilistically select pixels with a greater chance of impact.

5.2.8 Summary

This section has presented a modification to existing holographic search algorithms with a relative convergence error reduction improvement in the range of $[14.7\%, 19.2\%]$. Tests were run for direct search and simulated annealing algorithms as well as a range of test images, parameters and spatial light modulators with very similar performance improvements in all cases.

Unlike traditional search algorithms where test pixels are selected randomly, this section has presented a sorted variation where the pixels are sorted by the magnitude of the quantisation change immediately after back-projection. This was initially justified by some heuristic observations of the average nature of the back projected image and then trialled for a range of tests set-ups. While this is a general modification applicable to many HSAs and should not be regarded as a separate algorithm, it is proposed that this technique be referred to as sorted pixel selection (SPS).

While the performance improvement is small, it is consistent across a wide variety of test cases as well as being cheap and easy to apply to existing set-ups. For large images, search algorithms can take many hours to run and the improvements observed offer a significant benefit.

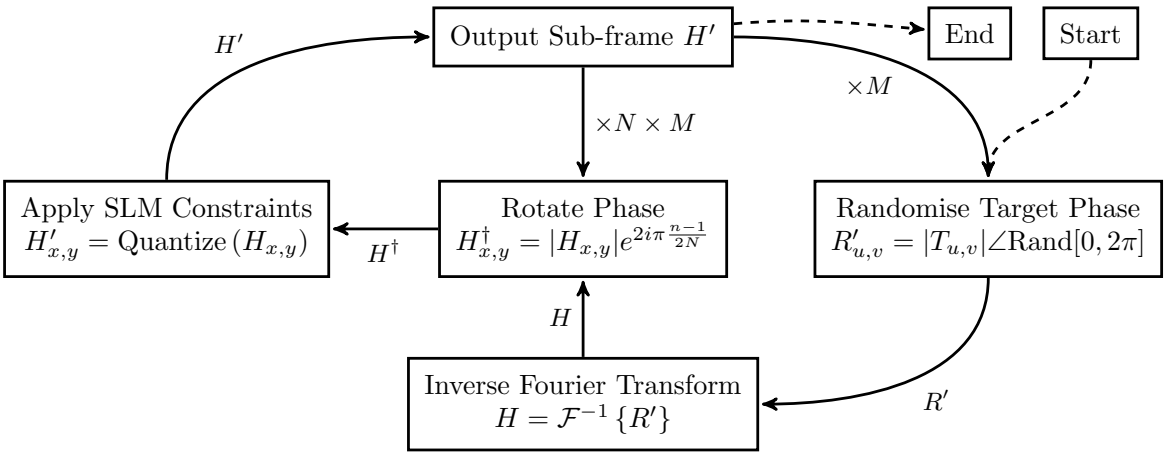


Fig. 5.10 Single-Transform Time-Multiplexed algorithm

5.3 Single-Transform Time-Multiplexed[2]

5.3.1 Introduction

In section 2.4.5, the OSPR algorithm was introduced, a computationally efficient hologram generation algorithm specifically designed with video display applications in mind and capable of generating subjectively pleasing holograms in real-time. The algorithm produces time-multiplexed sub-frames but is otherwise distinct from other holographic time-multiplexing techniques that, for example, aim to reduce speckle [281–283] or produce full-colour holograms [284–286]. In this section a significant improvement on the OSPR algorithm is presented, the single-transform time-multiplexed (STTM) algorithm.

5.3.2 Algorithm Summary

OSPR performs N independent DFT operations. This is computationally expensive in real-time applications where 1000s of sub-frames may require processing every second. Instead, an alternative algorithm is proposed shown in Figure 5.10. The phase of the target field

[2]The contents of this section were previously included as part of the following publication:

Peter J. Christopher, Ralf Mouthaan, Vamsee Bheemireddy, and Timothy D. Wilkinson, “Improving performance of single-pass real-time holographic projection", *Optics Communications* (2020), Volume 457, pp. 1456-1462, DOI: 10.1016/j.optcom.2019.124666

pattern is first randomised and the inverse DFT is taken to obtain the diffraction aperture, in a similar manner to many other algorithms. At this stage the complex values of the obtained hologram are rotated through a phase angle of $2\pi\frac{n-1}{N}$ before the SLM constraints are applied to obtain the sub-frame. This is repeated for $n = 1..N$. As the DFT magnitude is invariant under rotation in the complex plane, this does not effect the far-field amplitudes. This change in approach significantly reduces the computational load when compared to OSPR. It is noted that the STTM algorithm easily lends itself to parallel execution.

For values of N larger than 5, it is often preferred to periodically restart the algorithm. Consequently, the replay field is phase-randomised and the inverse DFT taken M times, and the phase angle of the hologram is shifted by $2\pi\frac{n-1}{N}$ for $n = 1..N$ times, to yield M sets of N sub-frames. This is discussed further in Section 5.3.5

5.3.3 Results

Figure 5.11 shows the phase-insensitive mean-squared error (MSE) convergence of OSPR as well as STTM and hybrid STTM, as calculated from Eq (5.4). This error metric is adopted as it encompasses both bias and variance errors. Values are taken as being the mean of 100 independent runs with error bars showing one standard deviation. The 512×512 pixel *Mandrill* test image with artificially induced rotational symmetry of Figure 5.11 (left) is used for the target.

A comparison of the computer-generated replay fields generated by OSPR, STTM and hybrid STTM is shown in Figure 5.12. The shown images correspond to the equally-weighted sum of the obtained sub-frames to mimic the impulse response of the human eye. Low resolutions are provided to ease comparison. The STTM image (centre right) had a mean-squared error less than 20% greater than the OSPR generated frames (centre right) and was generated in less than 10% of the time. The combined hybrid frame (far right) was generated from 3 sets of 4 STTM sub-frames and had an error 5% less than the OSPR equivalent and was generated in 30% of the time. The relative speed up becomes even more significant at higher resolutions where the FFT step takes up a greater percentage of the performance impact.

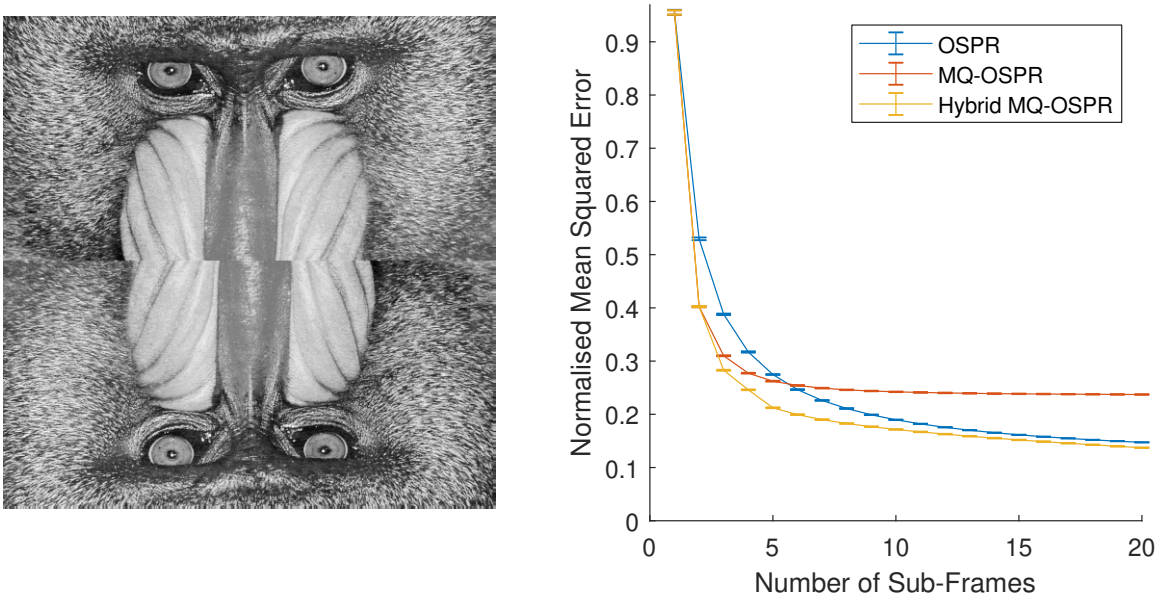


Fig. 5.11 Left: 512x512 Mandrill test image with induced rotational symmetry. Right: Time-averaged errors for different OSPR variants run on the Mandrill image (left). Values are taken as being the mean of 100 independent runs with error bars showing one standard deviation.

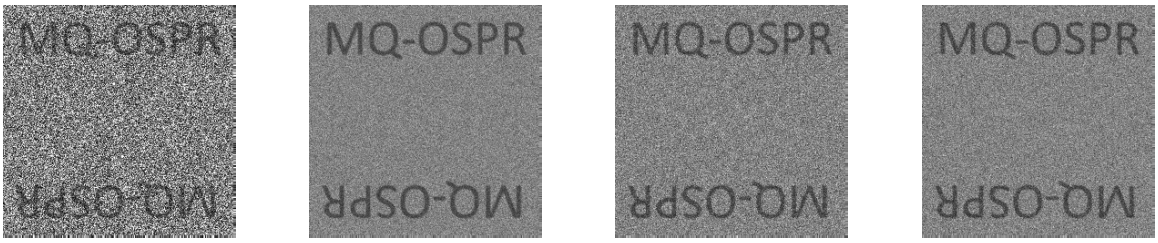


Fig. 5.12 Comparison of OSPR and STTM showing a single frame of OSPR (far left), 12 frames of OSPR (centre left), 12 frames of STTM (centre right) and 3 sets of 4 STTM sub-frames (far right). Target image is 256×256 pixels and the simulated device binary phase.

Fig. 5.13 Experimental implementation showing a single subframe (right) 25 OSPR subframes (centre) and 25 hybrid STTM subframes (right). Test image used is taken from the USC-SIPI database[49] and is shown in Figure 2.12. The hybrid STTM reproduction offers lower noise and improved contrast.

5.3.4 Experimental

Figure 5.13 shows a single subframe (right) 25 OSPR subframes (centre) and 25 hybrid STTM subframes (right) displayed using the experimental system described in Section 5.1.1 with only 512×512 pixels of the Jasper SLM being enabled to ensure a fair comparison. The Jasper is also operated in a binary manner and the subframes are shown are captured at defined points in the SLMs PWM update cycle before being linearly summed. The single frame shown is also summed over 25 exposures to reduce the effect of drift and other temporal issues. The hybrid STTM frames were generated in $< 20\%$ of the time taken to generate the OSPR frame.

5.3.5 Derivation

In order to provide a relationship for the expected MSE reduction a three-stage argument is made. Firstly, it is shown that the expected distribution of diffraction field magnitudes for any distribution of replay field magnitudes with uniformly distributed phase must follow a Rayleigh distribution. Secondly, it is shown that for a mean squared error (MSE) estimator the expected error of modifying a single pixel is proportional to the square of the distance moved. Thirdly, it is shown that using N binary quantised subframes are in fact equivalent to a single frame displayed on a device with $2N$ modulation levels. Finally these relationships are combined to provide an analytical relationship for error reduction against number of

iterations for STTM. This shows that the convergent error is expected to be $\approx 26\%$ of the first iteration error.

Expected distribution of diffraction field values

The first step followed is to develop a theory for the expected distribution of diffraction field values. If a replay field is considered with a distribution of amplitudes $\mathbb{R}_r$ and a distribution of phases Φ_r. $\mathbb{R}_r$ is assumed to be an arbitrary distribution with variance σ_r^2 which is here normalised to 1. Φ_r is assumed to be uniformly distributed in the interval $[0, 2\pi)$ and independent of $\mathbb{R}_r$.

$$\mathbb{C}_r = \mathbb{R}_r e^{i\Phi_r} \tag{5.1}$$

The diffraction field is related to the replay field by the inverse DFT. The distribution of values $\mathbb{C}_d$ taken on by a given diffraction field pixel is hence given by

$$\mathbb{C}_d = \frac{1}{\sqrt{N_u N_v}} \sum_{N_u} \sum_{N_v} \mathbb{R}_r e^{i\Phi_r} e^{2\pi i\left(\frac{ux}{N_u} + \frac{vy}{N_v}\right)} \tag{5.2}$$

where N_u and N_v here represent the number of pixels on the u and v axes respectively. This summation is over a set of vector variables and as such tends towards a Rayleigh distribution of the form of Eq (5.3) for large $N_u N_v$.

$$p(r) = 2re^{-r^2} \tag{5.3}$$

This result relies on the central limit theorem, and is consequently only valid for large $N_u N_v$. As the only other restriction placed on the replay field pixel magnitudes was that they be normalised to unit variance, this formula applies to any expected magnitude distribution in the replay field, not just a uniform distribution. For example, the expected diffraction field magnitudes and phases of the phase randomised 512×512 *Mandrill* test image is shown in Figure 5.3.

Expected error as a function of quantisation change

The per-pixel phase insensitive MSE formula is given as a function of the target image T and generated replay field R

$$E_{\text{MSE}}(T,R) = \frac{1}{N_x N_y} \sum_{u=0}^{v=N_x-1} \sum_{y=0}^{y=N_y-1} [|T_{u,v}| - |R_{u,v}|]^2 . \tag{5.4}$$

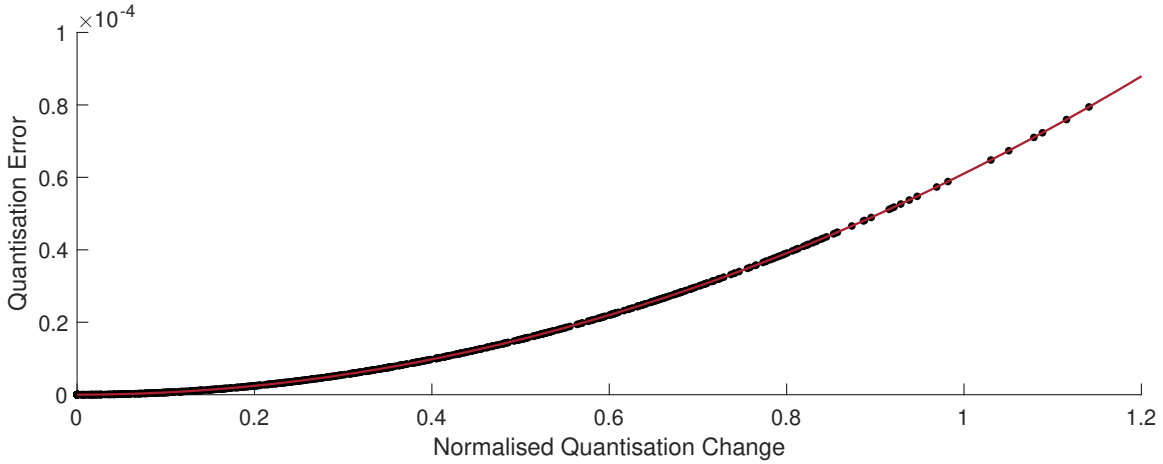


Fig. 5.14 Scatter plot of pixel value changes during quantisation of a continuous phase hologram against resultant error as well as the expected trend line.

The change in the a replay field pixel $\Delta R_{u,v}$ due to a change in a diffraction field pixel $\Delta Hx,y$ is derived from the DFT identity of Eq (2.72)

$$\Delta R_{u,v} = \frac{1}{\sqrt{N_x N_y}} \Delta H_{x,y} e^{\frac{ux}{N_x} + \frac{vy}{N_y}} \tag{5.5}$$

.

Inserting this into Eq (5.4) and performing the summation gives an expression for the change in MSE ΔE_{MSE} due to an altered hologram pixel where C_{MSE} is a constant in the range $[0,1)$. This relationship is also shown in Figure 5.14.

$$\Delta E_{MSE} = C_{MSE} \frac{|\Delta H_{x,y}|^2}{N_x N_y} \tag{5.6}$$

The constant C_{MSE} will decrease as further pixels are quantised and correlations are introduced between pixels. The analysis below depends only on the ratio between errors and the constant C_{MSE} will cancel.

Multi-frame equivalence to multi-level quantisation

By treating the time-multiplexed hologram-subframes as a linear addition of intensities it can be seen that summing the binary phase quantisation of a hologram and the binary phase quantisation of the same hologram rotated by 60° and 120° is the equivalent of the six phase quantisation of the hologram, as illustrated on the left side of Figure 5.6. More generally, the projection of N STTM frames quantised on an SLM with M levels is equivalent to projecting a single frame quantised on an SLM with NM levels. This leads in the limit as $N \to \infty$ to a continuously modulated phase device.

$$\sum_{n=1}^{N} \underset{\text{binary phase}}{\texttt{Quantise}} \left[He^{2i\pi\frac{n-1}{2N}} \right] = N \times \underset{2N\text{phase levels}}{\texttt{Quantise}} \left[He^{2i\pi\frac{1}{4N}} \right] \tag{5.7}$$

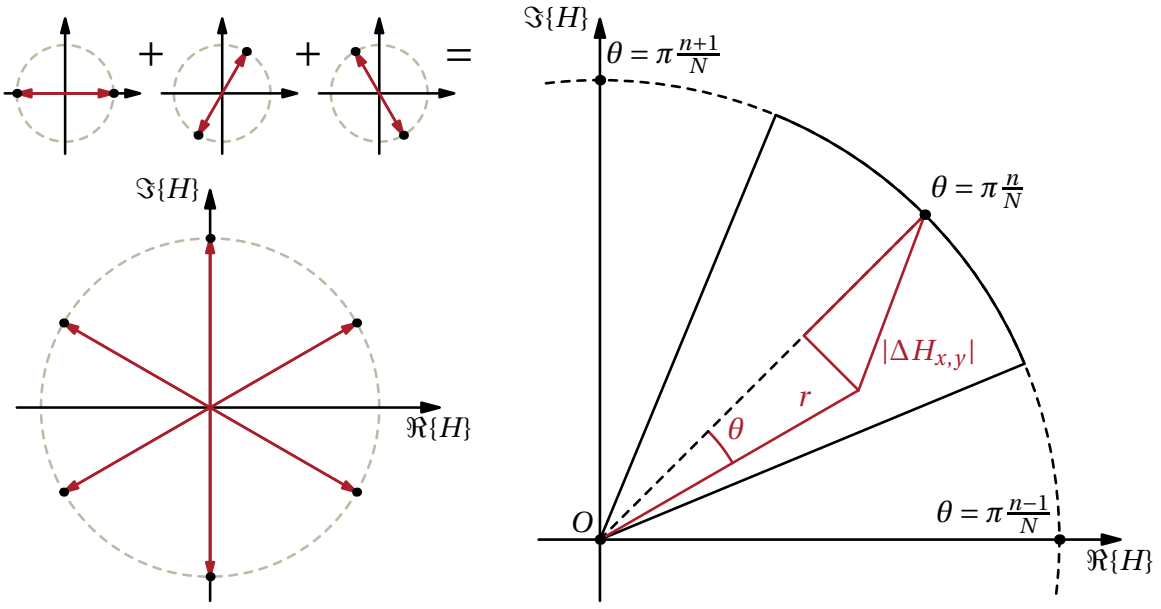


Fig. 5.15 Left: The sum of the binary phase quantisations of a hologram rotated through 0°, 60° and 120° is equivalent to a single hologram quantised on a six-level device. Right: Geometry of quantising and modulating a pixel.

Considering the display of N binary hologram subframes to be equivalent to a single subframe on an N-level devices allows any a given pixel value of the hologram $H_{x,y}$ can be written in

terms of amplitude r and angle θ from the nearest virtual modulation level where θ where θ is in the range $\left[-\frac{\pi}{2N}, \frac{\pi}{2N}\right)$ as shown in Figure 5.15 (right). The distance $\left|\Delta H_{x,y}\right|$ between $H_{x,y}$ and the nearest device level is then given by

$$\left|\Delta H_{x,y}\right| = \sqrt{(1 - r\cos\theta)^2 + (r\sin\theta)^2} = \sqrt{1 - 2r\cos\theta + r^2} \tag{5.8}$$

Combination

Section 5.3.5 gave the probability distributions of the magnitude and phase of the diffraction field. Assuming the distributions to be independent of each other, these can be combined with the error introduced by quantising a single pixel, calculated in Section 5.3.5, to give the expected value of the error due to quantising the entire hologram on an N-level phase modulator. It is noted that the integral is performed over the region of the argand diagram for which pixel values map onto a virtual modulation level - this treatment is justified given the conclusions drawn in Section 5.3.5.

$$E_{\text{MSE,tot}} = \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi/N} p(r)p(\theta)\Delta E_{MSE}(r,\theta)drd\theta \tag{5.9}$$

(5.3), (5.6) and (5.8) are then substituted in, and $p(\theta)$ is assumed to be a uniform distribution that integrates to unity. This then gives the following expression for the total error.

$$\begin{aligned} E_{\text{MSE,tot}} &= \frac{C_{\text{MSE}}}{N_x N_y} \int_0^{\infty} 2re^{-r^2} \int_{-\frac{\pi}{2N}}^{\frac{\pi}{2N}} \frac{1}{2\pi}(1 - 2r\cos\theta + r^2)d\theta dr \\ &= \frac{C_{\text{MSE}}}{N_x N_y}\left(2\pi - 2N\sqrt{\pi}\sin\left(\frac{\pi}{2N}\right)\right) \end{aligned} \tag{5.10}$$

Normalising to the single-frame case and considering the limit as $N \to \infty$

$$E_{\text{MSE,tot},\infty} = E_{\text{MSE,tot},1}\frac{\pi - \frac{1}{2}\pi^{\frac{3}{2}}}{\pi - \sqrt{\pi}} \approx 0.2611 E_{\text{MSE,tot},1} \tag{5.11}$$

which is in agreement with the observed behaviour.

5.3.6 Summary

The STTM hologram generation algorithm has been introduced, which allows subjectively pleasing time-multiplexed holograms to be generated in real-time. Generation times have been shown to be an order of magnitude faster than competing time-multiplex algorithms such as OSPR, with an associated degradation in performance of less than 20%. A mathematical rationale has been given for the performance of the STTM algorithm and a hybrid STTM/OSPR algorithm has been developed that combines the advantages of each of the two approaches. The speed-up offered by the STTM algorithm offers the potential for higher-resolution, higher-framerate and more cost effective holographic displays.

5.4 Holographic Predictive Search[3]

5.4.1 Introduction

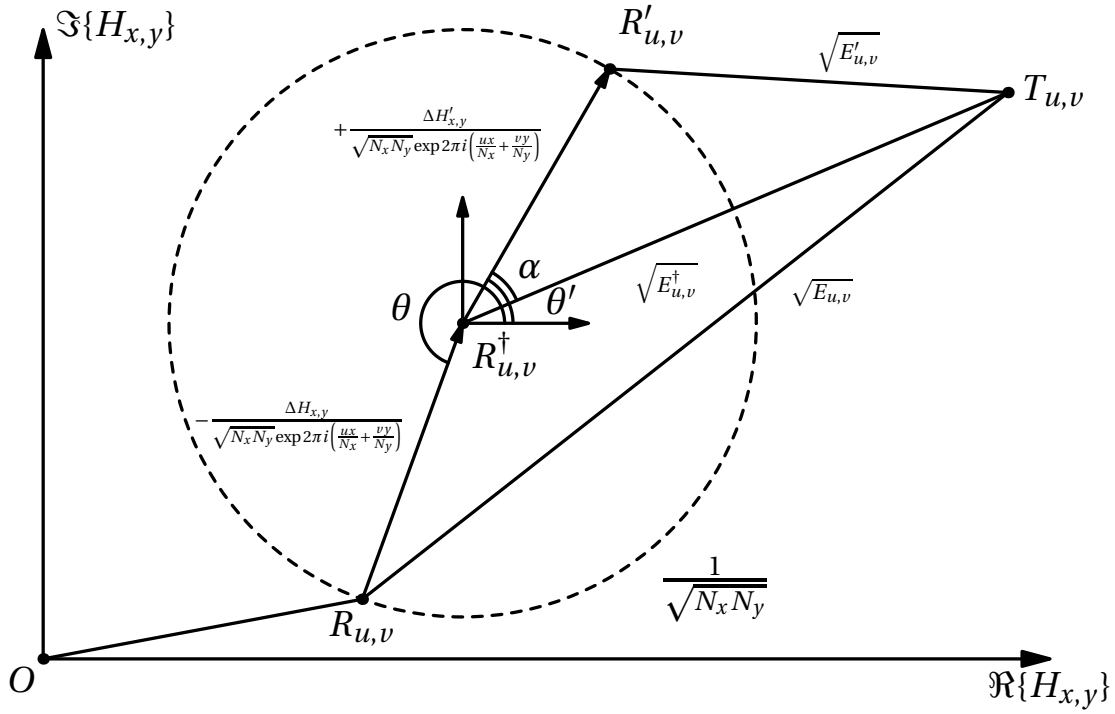
This section presents a new algorithm for computer-generated holography called holographic predictive search (HPS). HPS uses a prescient model of the Fourier transform to improve on the most common HSAs discussed in Sections 2.4.4 and 11.

5.4.2 Predictive Search

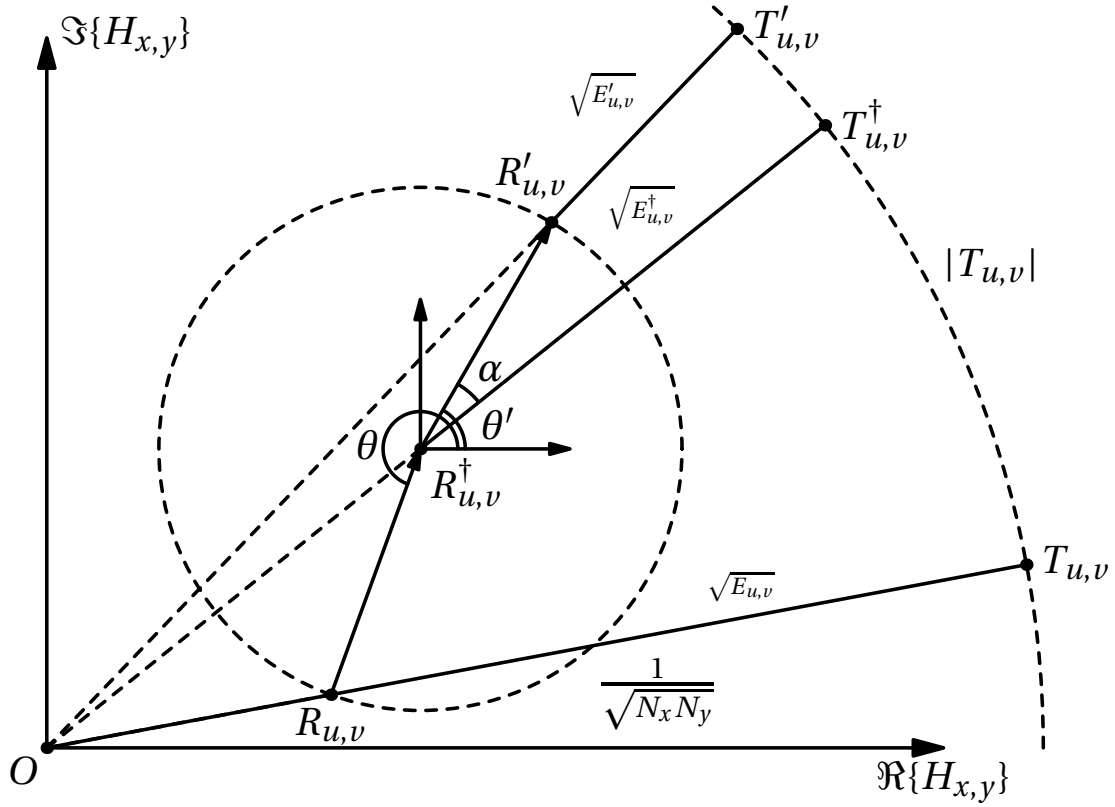
The goal is a prescient or predictive model for search algorithms. Randomly modifying a pixel and testing its effect on the replay field works well for binary holograms but has relatively poor performance for high numbers of modulation levels. Instead, a geometrical understanding of the update step in (2.72) can be used to derive a predictive relationship for the *best* new pixel value. The approach for this can be thought of as setting an individual

[3]The contents of this section were previously included as part of the following publications:

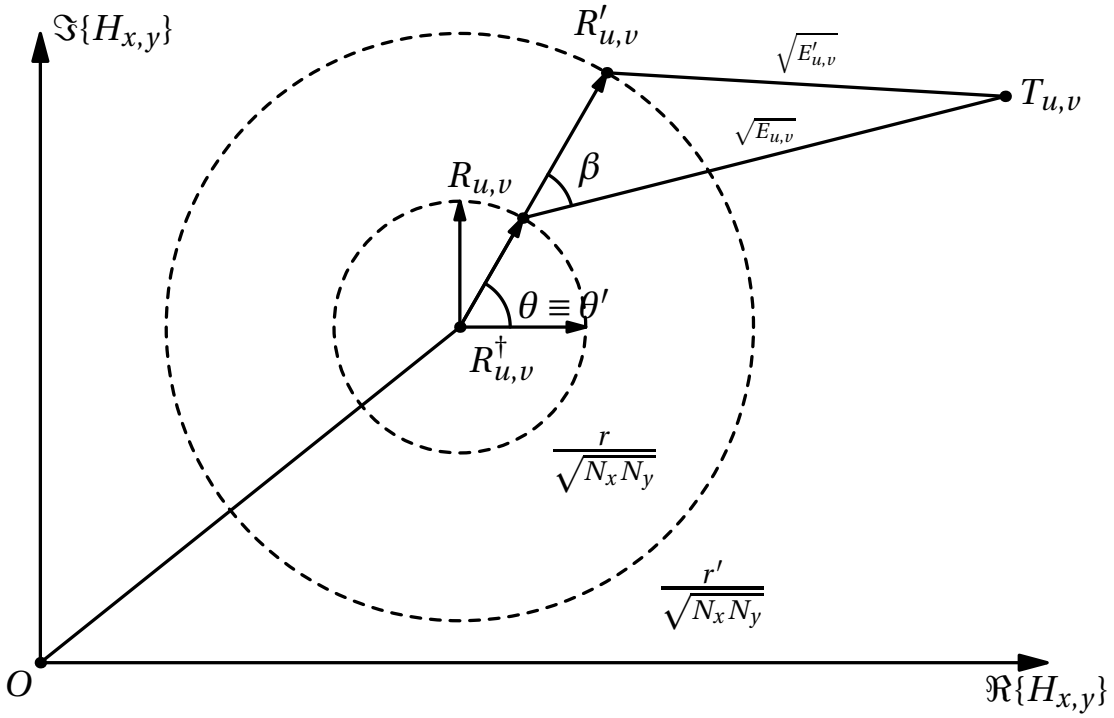
> Peter J. Christopher, Youchao Wang, and Timothy D. Wilkinson, “Predictive search algorithm for phase holography", *Journal of the Optical Society of America A* (2019), Volume 36, Issue 12, pp. 2068-2075, DOI: 10.1364/JOSAA.36.002068 Peter J. Christopher, Ralf Mouthaan, George S. D. Gordon and Timothy D. Wilkinson, “Holographic Predictive Search: Extending the Scope", *Optics Communications*, (2020), Volume 467,
> DOI: 10.1016/j.optcom.2020.125701



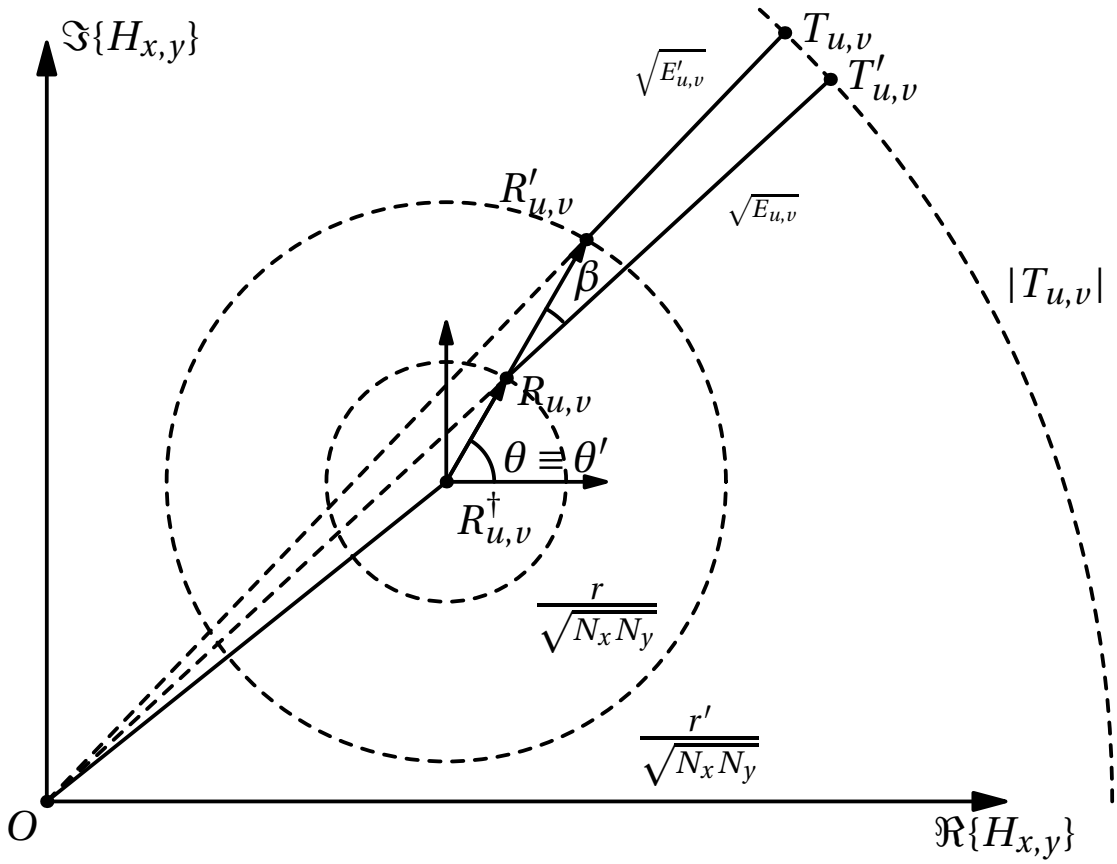
(a) Phase modulated, phase sensitive



(b) Phase modulated, phase insensitive



(c) Amplitude modulated, phase sensitive



(d) Amplitude modulated, phase insensitive

Fig. 5.16 Holographic predictive search problem geometry

pixel x, y to zero and then performing a relationship of the new error E' as a function of new phase angle θ'. Provided this relationship is linear, analytical techniques can be used to derive a relationship for the ideal value.

Scope

Unfortunately, there are three different distinctions that must be considered when choosing an appropriate algorithm.

- **Transform Type** - Only far-field or Fraunhofer holograms can be modelled as an Fourier Transform. For mid-field holograms, an additional quadratic phase term must be added to form a Fresnel Transform.
- **SLM Modulation Behaviour** - SLMs typically modulate in either phase or amplitude. Each paradigm involves a different set of relationships between the SLM and the replay field.
- **Phase Sensitivity** - Many display applications, however, do not require the phase constraint due to the human eye's phase insensitivity. This additional freedom requires a separate formulation.

To give an understanding of the importance of modulation behaviour and phase sensitivity, Figures 5.17 and 5.18 take an initial inverse Fourier transform of the *Mandrill* test image and plot the effect on mean squared error (MSE) of changing the level on 10 randomly selected SLM pixels for two different categories of SLM - phase and amplitude modulating - and for two different categories of replay field - phase sensitive and insensitive. These were generated by selecting a random hologram pixel $H_{x,y}$ and plotting the change in final error for a range of pixel values. They show that the response to level changes of a single phase pixel has a near sinusoidal effect on the error of the replay field whereas changing an amplitude pixel has a more linear response. The SLM is assumed to be 256×256 pixels with a flat unit intensity illumination.

The aim of HPS is to mathematically capture the nature of the curves shown and solve for the optimum location without having to apply (2.72) to every step. In practice the phase insensitive cases degenerate to quartic polynomials and it is only through the judicious use of simplifications that high speed solutions can be achieved.

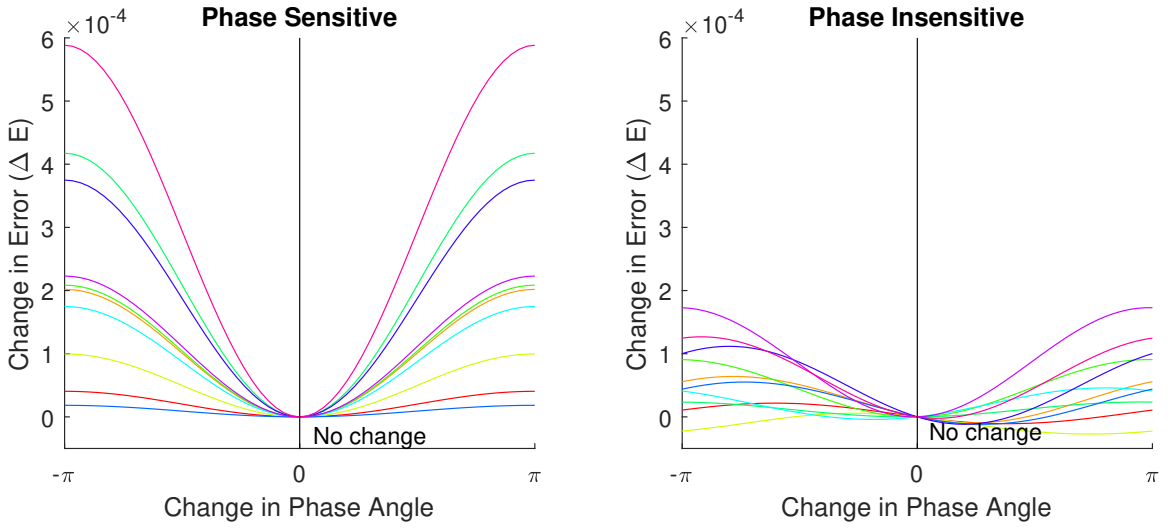


Fig. 5.17 Selection of final errors depending on pixel changes for phase modulating SLMs for phase sensitive replay fields (left) and phase insensitive replay fields (right).

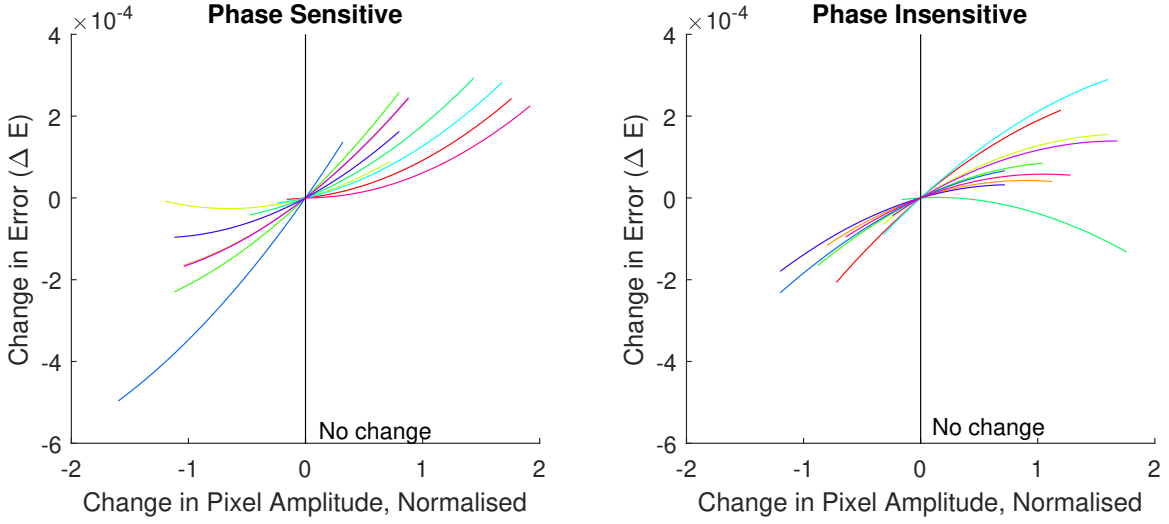
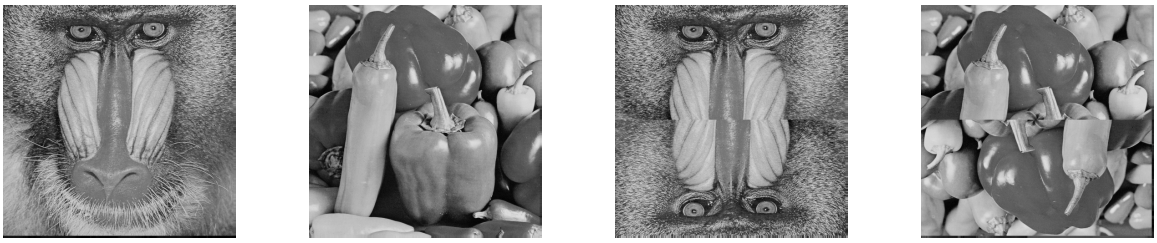


Fig. 5.18 Selection of final errors depending on pixel changes for amplitude modulating SLMs for phase sensitive replay fields (left) and phase insensitive replay fields (right).



(a) (b) (c) (d)

Fig. 5.19 The two test images used showing the *Mandrill* and *Peppers* as well as their artificially symmetric counterparts.

The necessary background to conform to every combinations of these constraints is presented in the sections to come.

Methods

It is challenging to fairly compare the performance of techniques across different system designs. In order to aid comparison, the following conventions have been established:

1. The *Mandrill* test image shown in Figure 5.19a is used to provide the target intensities.
2. Amplitude holograms are generated with the rotationally symmetric version of Figure 5.19c to avoid error due to image symmetry.
3. Phase sensitive holograms are generated with the *Peppers* test image used used as the phase component as shown in Figure 5.19b with rotationally symmetric variant shown in Figure 5.19d.
4. For phase insensitive holograms, the entire target region is solved for. In the case of phase sensitive holograms, for reasons of degrees of freedom, the target image is scaled to only fill the central quadrant of the initial replay field and the surrounding regions set to zero.
5. Planar unit intensity incident on the hologram is assumed with the target scaled to ensure conservation of energy.

These differences in method mean that the normalised error metrics used should be treated as distinct in each case and cannot be compared quantitatively between cases. The Fraunhofer cases are treated first in Section 5.4.3-5.4.6.

5.4.3 Phase Modulated SLM, Phase Sensitive Replay Field

The first Fraunhofer case to consider is that of a phase modulated SLM with a phase sensitive replay field.

Derivation

Setting an individual diffraction field pixel x,y to zero will introduce an error into each location u,v in the replay field R given by (2.72) with $\Delta H_{x,y} = -H_{x,y}$ leading to a modified replay field $R^{\dagger}$. Figure 5.16a models this geometrically on the Argand diagram.

The task is to find a $H'_{x,y}$ of unit magnitude such that the error across the new replay field R' is minimised. Expressing θ and θ' - the respective old and new pixel phase angles - in terms of unknown $\angle H'_{x,y}$ and known diffraction field coordinates x, y; replay field coordinates u, v and resolutions N_x, N_y allows the problem to be treated trigonometrically.

$$\theta = \angle H_{x,y} - 2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right), \quad \theta' = \angle H'_{x,y} - 2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) \tag{5.12}$$

Note that $\angle X$ here refers to the phase angle of X.

The error after zeroing pixel x,y is given as $E^{\dagger}_{u,v} = \left|T_{u,v} - R^{\dagger}_{u,v}\right|^2$ which is knowable at runtime. The new error $E'_{u,v}$ is given as a function of α

$$\begin{aligned} E'_{u,v} &= \left|T_{u,v} - R'_{u,v}\right|^2 \\ &= \left[\left|T_{u,v} - R^{\dagger}_{u,v}\right| - \frac{\cos\alpha}{\sqrt{N_x N_y}}\right]^2 + \left[\frac{\sin\alpha}{\sqrt{N_x N_y}}\right]^2 \\ &= E^{\dagger}_{u,v} + \frac{\cos^2\alpha}{N_x N_y} - 2\sqrt{E^{\dagger}_{u,v}}\frac{\cos\alpha}{\sqrt{N_x N_y}} + \frac{\sin^2\alpha}{N_x N_y} \end{aligned} \tag{5.13}$$

Since $\cos^2\alpha + \sin^2\alpha = 1$, the change in error for any given α is

$$\begin{aligned}\Delta E'_{u,v} &= E'_{u,v} - E^{\dagger}_{u,v} \\ &= \frac{1}{N_x N_y} - 2\sqrt{E^{\dagger}_{u,v}}\frac{\cos\alpha}{\sqrt{N_x N_y}},\end{aligned} \tag{5.14}$$

where α is given from θ'

$$\alpha = \theta' - \angle(T_{u,v} - R^{\dagger}_{u,v}) = \angle H'_{x,y} - \left[2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) + \angle(T_{u,v} - R^{\dagger}_{u,v})\right] \tag{5.15}$$

Using $\cos a - b = \cos a \cos b + \sin a \sin b$ and substituting into (5.14)

$$\Delta E'_{u,v} = \frac{1}{N_x N_y} - 2\sqrt{\frac{E^{\dagger}_{u,v}}{N_x N_y}}\left[\cos\theta_{H'}\cos C_{u,v} + \sin\theta_{H'}\sin C_{u,v}\right]$$

$$\begin{aligned}\textit{where} \quad \theta_{H'} &= \angle H'_{x,y} \\ C_{u,v} &= 2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) + \angle(T_{u,v} - R^{\dagger}_{u,v})\end{aligned} \tag{5.16}$$

Summing $\Delta E'_{u,v}$ in both dimensions,

$$\begin{aligned}\Delta E' &= \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\Delta E'_{u,v} \\ &= 1 - \frac{2}{\sqrt{N_x N_y}}\left[\cos\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E^{\dagger}_{u,v}}\cos C_{u,v} + \sin\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E^{\dagger}_{u,v}}\sin C_{u,v}\right].\end{aligned} \tag{5.17}$$

Taking ${}^{\mathrm{d}\Delta E'}/_{\mathrm{d}\theta_{H'}} = 0$ to find the the value of $\theta_{H'}$ where $\Delta E'$ is minimum

$$\sin\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E^{\dagger}_{u,v}}\cos C_{u,v} - \cos\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E^{\dagger}_{u,v}}\sin C_{u,v} = 0 \tag{5.18}$$

which is trivially solvable

$$\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\sin C_{u,v}\right)}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\cos C_{u,v}\right)}\right] \tag{5.19}$$

The correct solution is given using $\mathrm{d}^2\Delta E'/\mathrm{d}\theta^2_{H'} > 0$

$$\cos\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E^{\dagger}_{u,v}}\cos C_{u,v} + \sin\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E^{\dagger}_{u,v}}\sin C_{u,v} > 0 \tag{5.20}$$

Performance

This result allows the algorithm to more than trial a new pixel phase as in DS and SA algorithms. Instead it is possible to use a known relationship to determine the best possible phase for that pixel. The cost of this is an increased overhead on each iteration.

In the binary modulation case this technique offers no benefit but when applied to the multi-phase or continuous-phase devices it can significantly reduce the required number of iterations as it will find the best possible pixel phase rather than checking one alternative value.

As there are no approximations used in this derivation, and instead rely solely on the linearity of adding frequency components, this approach is guaranteed to analytically find the best value for a given pixel. Unfortunately, this does not remove the risk of local minima.

Putting this in algorithmic form leads to Algorithm 5.1 Note that this ignores the choice of solutions due to the $\tan^{-1}$ element. Most computer implementations of $\tan^{-1}$ return the value for $\theta_{H'}$ in the range $[-\pi/2, \pi/2)$. Simply treat $\theta_{H'} \leftarrow \theta_{H'} + \pi$ if (5.20) does not hold.

HPS can be compared to traditional HSAs. Using the target images from Section 5.4.2 on a 2^8 level phase SLM gives the performance graph as shown in Figure 5.20 where DS is shown in blue and HPS in orange.

Figure 5.20 shows the case for a phase sensitive problem where only the central quadrant is taken as the region of interest with regions outside being set to zero target energy. This gives an approximately $10\times$ improvement in convergence time over $1,000,000$ iterations though this number varies dependant on other factors. Of note is that while the computation load of

Algorithm 5.1: Phase Modulated, Phase Sensitive HPS

Back-propagate the target to the diffraction plane: $H_{u,v} = \mathscr{F}^{-1}\{R'\}_{u,v}$

Quantise the resultant hologram: $H' = \text{quantise}(H_{u,v})$

Generate initial replay field: $R_{u,v} = \mathscr{F}\{H'\}_{u,v}$

Generate initial error: $E_{u,v} = |T_{u,v} - R_{u,v}|^2$

for $n \leftarrow 1$ **to** N **do**

Zero a random pixel x, y: $R^{\dagger}_{u,v} = R_{u,v} - \frac{H_{x,y}\exp\left[-2\pi i\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)\right]}{\sqrt{N_x N_y}}$

Calculate modified error: $E^{\dagger}_{u,v} = \left|T_{u,v} - R^{\dagger}_{u,v}\right|^2$

Calculate constant: $C_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right) + \angle(T_{u,v} - R^{\dagger}_{u,v})$

Calculate new pixel phase angle: $\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\cos C_{u,v}\right)}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\sin C_{u,v}\right)}\right]$

Calculate new pixel: $H'_{x,y} = e^{i\theta_{H'}}$

Calculate new replay field: $R'_{u,v} = \frac{\left(H'_{x,y} - H_{x,y}\right)\exp\left[-2\pi i\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)\right]}{\sqrt{N_x N_y}}$

Calculate new error: $E'_{u,v} = \left|T_{u,v} - R'_{u,v}\right|^2$

end

a single iteration is higher, HPS is mathematically guaranteed to at least match DS in terms of performance per iteration.

Here the final error for HPS is less than 10% of that of DS and can take up to $10\times$ fewer iterations to reach a given target error. Efficiency is, mathematically, very high with $\gg 99\%$ of the energy being contained in the central quadrant.

Very similar performance improvements were seen when used within a simulated annealing algorithm. Using the same test configuration, HPS outperformed SA by up to $10\times$ in terms of quality after a given number of iterations.

In order to visually understand the performance improvement the 256×256 *Mandrill* and *Peppers* test images shown in Figure 5.19 were encoded into the central quadrant of a 512×512 target as the amplitude and phase terms respectively. The results of running $1,000,000$ iterations of DS and HPS are shown in Figure 5.21.

It will be seen from this figure that the image quality is good in both cases with HPS being visually superior to DS. The SSIM values given are calculated with a dynamic range of 1.0.

5.4.4 Phase Modulated SLM, Phase Insensitive Replay Field

The phase insensitivity of the eye means that display applications are often phase insensitive. This greatly increases the problem freedom but also changes the predictive geometry into a non-linear problem. The updated regime is shown on the Argand diagram in Figure 5.16b.

Derivation

Zeroing an individual SLM pixel x,y introduces error to location u,v in the replay field R given by (2.72) with $\Delta H_{x,y} = -H_{x,y}$. This replay field we term $R^{\dagger}$. The task is to find new pixel value $H'_{x,y}$ of unit magnitude so that the error in the new replay field R' is minimised. Expressing θ and θ', the old and new pixel phases, in terms of unknown $\angle H'_{x,y}$ and known x, y, u, v, N_x and N_y,

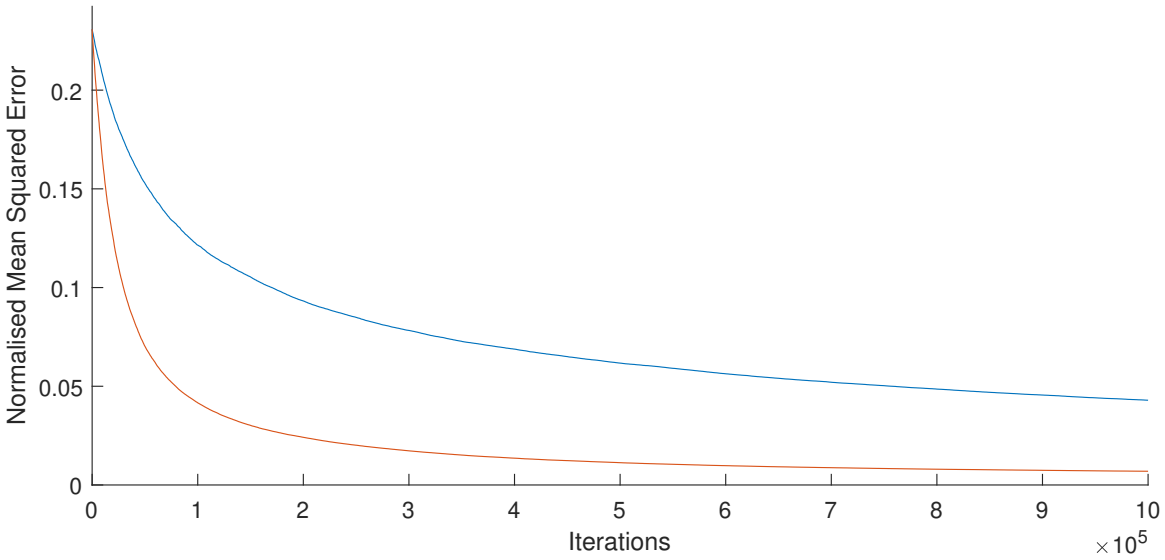


Fig. 5.20 Comparison of Direct Search (blue) against Phase Sensitive Predictive Holographic Search (orange).

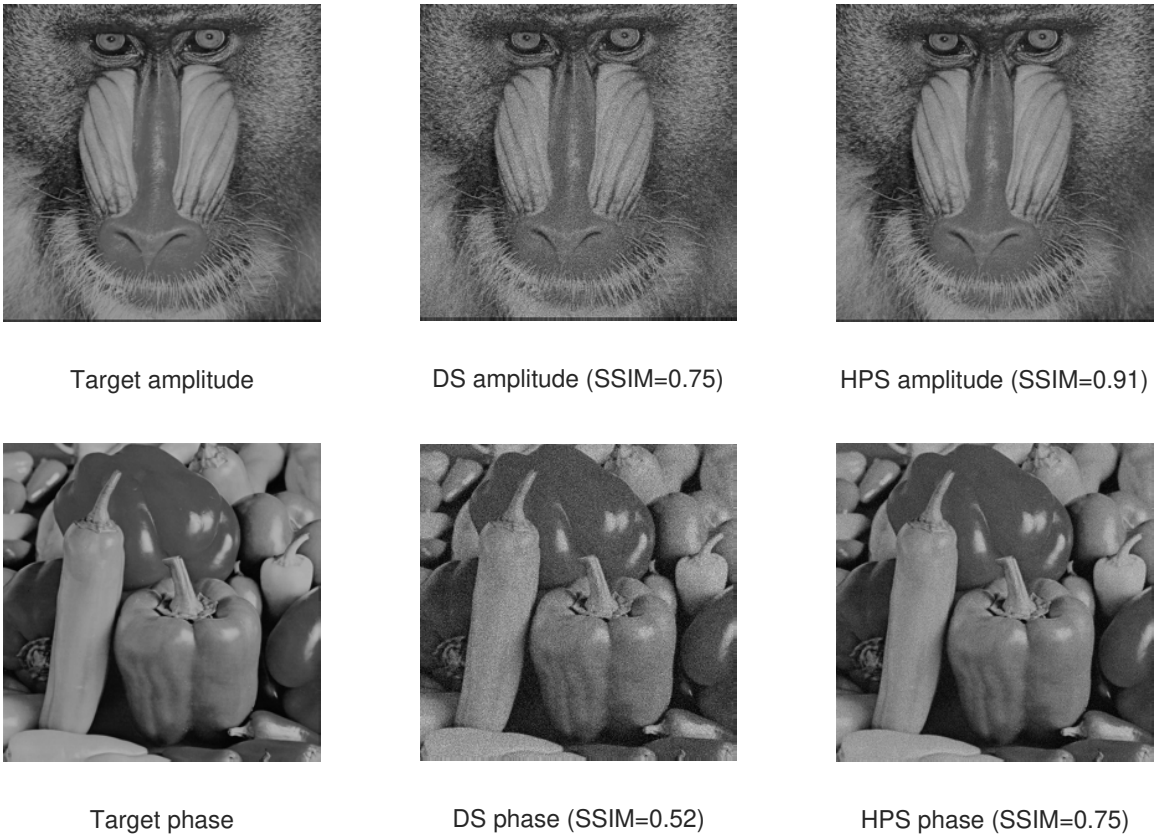


Fig. 5.21 Comparison of Direct Search (centre) against Phase Sensitive Predictive Holographic Search (right).

$$\theta = \angle H_{x,y} - 2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right),$$
$$\theta' = \angle H'_{x,y} - 2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) \tag{5.21}$$

The error after zeroing pixel x,y is given as $E^{\dagger}_{u,v} = \left(|T_{u,v}| - \left|R^{\dagger}_{u,v}\right|\right)^2$. The new error $E'_{u,v}$ is given by

$$\begin{aligned}
E'_{u,v} &= \left[|T_{u,v}| - \left|R'_{u,v}\right|\right]^2 \\
&= \left[|T_{u,v}| - \sqrt{\left|R^{\dagger}_{u,v}\right|^2 + \frac{1}{N_xN_y} + \frac{2\left|R^{\dagger}_{u,v}\right|\cos\alpha}{\sqrt{N_xN_y}}}\right]^2 \\
\Rightarrow \Delta E'_{u,v} &= E'_{u,v} - E^{\dagger}_{u,v} \\
&= \frac{1}{N_xN_y} + \frac{2\left|R^{\dagger}_{u,v}\right|\cos\alpha}{\sqrt{N_xN_y}} + 2|T_{u,v}|\left|R^{\dagger}_{u,v}\right| \\
&\quad - 2|T_{u,v}|\sqrt{\left|R^{\dagger}_{u,v}\right|^2 + \frac{1}{N_xN_y} + \frac{2\left|R^{\dagger}_{u,v}\right|\cos\alpha}{\sqrt{N_xN_y}}}
\end{aligned} \tag{5.22}$$

As

$$\alpha = \theta' - \angle R^{\dagger}_{u,v} = \angle H'_{x,y} - \left[2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) + \angle R^{\dagger}_{u,v}\right] \tag{5.23}$$

applying the Taylor expansion of $\sqrt{1+z}$ to give

$$\Delta E'_{u,v} = D_{u,v} + F_{u,v}\left[\cos\theta_{H'}\cos C_{u,v} + \sin\theta_{H'}\sin C_{u,v}\right]$$

where

$$
\begin{aligned}
D_{u,v} &= \frac{1}{N_x N_y} + 2|T_{u,v}|\left|R^{\dagger}_{u,v}\right| - 2|T_{u,v}|\sqrt{\left|R^{\dagger}_{u,v}\right|^2 + \frac{1}{N_x N_y}} \\
F_{u,v} &= \frac{2\left|R^{\dagger}_{u,v}\right|}{\sqrt{N_x N_y}} - \frac{2|T_{u,v}|\left|R^{\dagger}_{u,v}\right|}{\sqrt{N_x N_y}\sqrt{\left|R^{\dagger}_{u,v}\right|^2 + \frac{1}{N_x N_y}}} \\
C_{u,v} &= 2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) + \angle R^{\dagger}_{u,v}
\end{aligned} \tag{5.24}
$$

Summing over all pixels

$$
\begin{aligned}
\Delta E' &= \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} \Delta E'_{u,v} \\
&= \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} D_{u,v} + \cos\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} F_{u,v}\cos C_{u,v} + \sin\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} F_{u,v}\sin C_{u,v}
\end{aligned} \tag{5.25}
$$

Taking ${}^{\mathrm{d}\Delta E'}/_{\mathrm{d}\theta_{H'}} = 0$ to find the the value of $\theta_{H'}$ where $\Delta E'$ is minimum

$$
\sin\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} F_{u,v}\cos C_{u,v} - \cos\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} F_{u,v}\sin C_{u,v} = 0 \tag{5.26}
$$

which is trivially solvable.

$$
\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(F_{u,v}\sin C_{u,v}\right)}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(F_{u,v}\cos C_{u,v}\right)}\right] \tag{5.27}
$$

The correct solution can be chosen using ${}^{\mathrm{d}^2\Delta E'}/_{\mathrm{d}\theta^2_{H'}} > 0$

$$
\cos\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} F_{u,v}\cos C_{u,v} + \sin\theta_{H'}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} F_{u,v}\sin C_{u,v} > 0 \tag{5.28}
$$

This translates into Algorithm 5.2

Algorithm 5.2: Phase Modulated, Phase Insensitive HPS

Input: Target $T_{u,v}$, current replay $R_{u,v}$ and number of iterations N

Randomise target image phase: $R'_{u,v} = |T_{u,v}| \angle \text{Rand}[0, 2\pi]$

Back-propagate the target to the diffraction plane: $H_{u,v} = \mathscr{F}^{-1}\{R'\}_{u,v}$

Quantise the resultant hologram: $H' = \text{quantise}(H_{u,v})$

Generate initial replay field: $R_{u,v} = \mathscr{F}\{H'\}_{u,v}$

Generate initial error: $E_{u,v} = \left|T_{u,v}^2 - R_{u,v}^2\right|$

for $n \leftarrow 1$ **to** N **do**

Zero a random pixel x, y: $R^{\dagger}_{u,v} = R_{u,v} - \frac{H_{x,y} \exp\left[-2\pi i\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right)\right]}{\sqrt{N_x N_y}}$

Calculate modified error: $E^{\dagger}_{u,v} = \left(|T_{u,v}| - \left|R^{\dagger}_{u,v}\right|\right)^2$

Calculate constant: $C_{u,v} = 2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) + \angle R^{\dagger}_{u,v}$

Calculate constant: $F_{u,v} = \frac{2\left|R^{\dagger}_{u,v}\right|}{\sqrt{N_x N_y}} - \frac{2|T_{u,v}|\left|R^{\dagger}_{u,v}\right|}{\sqrt{N_x N_y}\sqrt{\left|R^{\dagger}_{u,v}\right|^2 + \frac{1}{N_x N_y}}}$

Calculate new pixel phase angle: $\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}(F_{u,v}\cos C_{u,v})}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}(F_{u,v}\sin C_{u,v})}\right]$

Calculate new pixel: $H'_{x,y} = e^{i\theta_{H'}}$

Calculate new replay field: $R'_{u,v} = \frac{\left(H'_{x,y} - H_{x,y}\right)\exp\left[-2\pi i\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right)\right]}{\sqrt{N_x N_y}}$

Calculate new error: $E'_{u,v} = \left(|T_{u,v}| - \left|R'_{u,v}\right|\right)^2$

end

Performance

Using the test regime from Section 5.4.2 on a 2^8 level phase SLM gives the performance graph shown in Figure 5.22.

Comparison with the DS case presents a significant improvement in convergence speed but this is less marked than the phase sensitive case. This is not unexpected as an examination of Figure 5.17 will show that the degree of variation per pixel is much lower in the PI case. i.e. the hologram initial hologram is much closer to the theoretical *best* hologram than in the PS case. Nonetheless, the error at the end of $100,000$ iterations is $10\times$ lower for HPS than for DS.

Figure 5.22 also provides a comparison of the 2^8 modulation level (left) vs 2^4 modulation level (right) cases. This shows a similar performance improvement in both cases. The effect of number of modulation levels is returned to later, but for now it should be noted that the primary difference is in an increased convergent error with other features remaining indistinguishable.

The algorithm reconstruction is shown in Figure 5.23 showing target image (left), DS (centre) and HPS (right). The SSIM values given are calculated with a dynamic range of 1.0.

Note also the use of the Taylor series expansion. An exact solution quickly devolves into a quartic polynomial which, while solvable, proves expensive computationally. Note also that it would be entirely possible to use more terms of the Taylor series for increased accuracy at the cost of performance. Initial tests suggested that the use of only two terms was sufficient with error in predicted angle $\theta_{H'}$ never going above 1%.

Finally, it is noted that while HPS compares favourably with DS in this case, it is unlikely to offer benefits over Gerchberg-Saxton for high numbers of modulation levels [50]. This discussion is returned to later.

5.4.5 Amplitude Modulated SLM, Phase Sensitive Replay Field

The third Fraunhofer case to consider is that of a amplitude modulated SLM with a phase sensitive replay field.

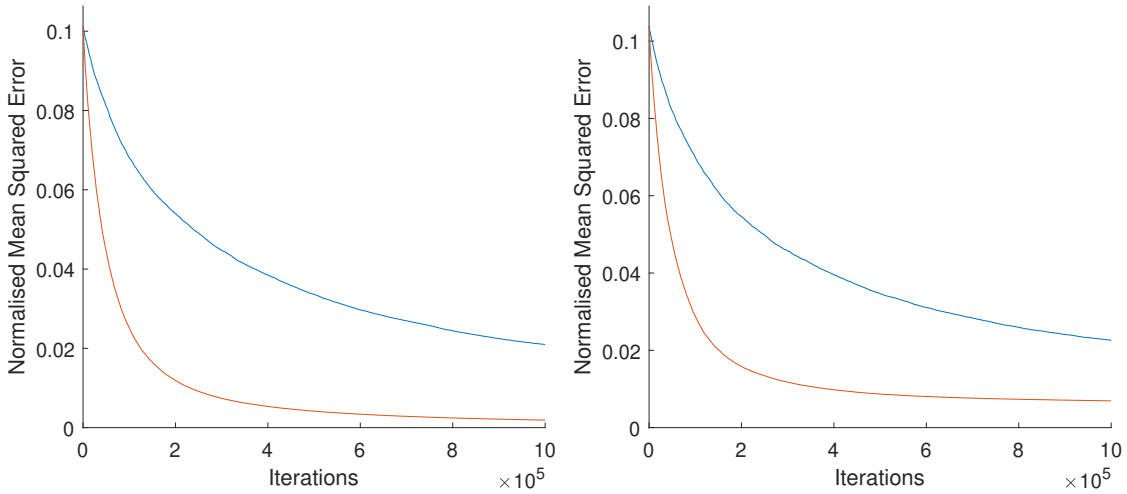


Fig. 5.22 Comparison of Direct Search (blue) against phase modulated, phase insensitive HPS (orange).

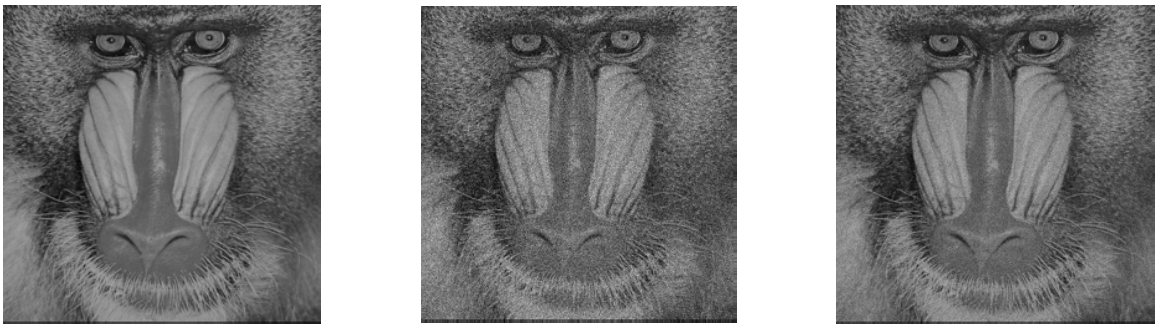


Fig. 5.23 Comparison of Direct Search (centre) against phase modulated, phase insensitive HPS (right).

Derivation

Similar to its phase modulated counterpart, amplitude modulated phase sensitive HPS has a problem geometry as shown in Figure 5.16c. Unlike in the phase modulated case, the new pixel phase angle $\theta_{H'}$ is equal to zero. As a result, the pixel zeroing step $R^{\dagger}$ and $E^{\dagger}$ can be ignored. Instead work is conducted in terms of $\Delta r' = r' - r$ where r and r' are the old and new pixel magnitudes respectively.

The initial error before modifying pixel x, y is given as $E_{u,v} = |T_{u,v} - R_{u,v}|^2$ which is knowable at runtime. The new error $E'_{u,v}$ is given by

$$\begin{aligned} E'_{u,v} &= \left|T_{u,v} - R'_{u,v}\right|^2 \\ &= \left[|T_{u,v} - R_{u,v}| - \frac{\Delta r}{\sqrt{N_x N_y}}\cos\beta\right]^2 + \left[\frac{\Delta r}{\sqrt{N_x N_y}}\sin\beta\right]^2 \\ &= E_{u,v} + \frac{\Delta r^2}{N_x N_y}\cos^2\beta - 2\sqrt{E_{u,v}}\frac{\Delta r}{\sqrt{N_x N_y}}\cos\beta + \frac{\Delta r^2}{N_x N_y}\sin^2\beta \end{aligned} \tag{5.29}$$

Remembering that $\cos^2\beta + \sin^2\beta = 1$, the change in error for given $\frac{\Delta r}{\sqrt{N_x N_y}}$ is

$$\begin{aligned} \Delta E'_{u,v} &= E'_{u,v} - E_{u,v} \\ &= \frac{\Delta r^2}{N_x N_y} - 2\sqrt{E_{u,v}}\frac{\Delta r}{\sqrt{N_x N_y}}\cos\beta \end{aligned} \tag{5.30}$$

$\beta_{u,v}$ is given from $\theta_{u,v}$

$$\beta_{u,v} = \theta_{u,v} - \angle(T_{u,v} - R_{u,v}) = -2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) - \angle(T_{u,v} - R_{u,v}) \tag{5.31}$$

where

$$\theta' = \theta = -2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) \tag{5.32}$$

Summing

Algorithm 5.3: Amplitude Modulated, Phase Sensitive HPS

1 Back-propagate the target to the diffraction plane: $H_{u,v} = \mathscr{F}^{-1}\{R'\}_{u,v}$

2 Quantise the resultant hologram: $H' = \text{quantise}(H_{u,v})$

3 Generate initial replay field: $R_{u,v} = \mathscr{F}\{H'\}_{u,v}$

4 Generate initial error: $E_{u,v} = |T_{u,v} - R_{u,v}|^2$

for $n \leftarrow 1$ **to** N **do**

5 Select a random pixel x, y

6 Calculate constant: $\beta_{u,v} = -2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) - \angle(T_{u,v} - R_{u,v})$

7 Calculate new pixel magnitude: $\Delta r = \frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E_{u,v}}\cos\beta}{\sqrt{N_xN_y}}$

8 Calculate new pixel value: $H'_{x,y} = \min(\max(H'_{x,y} + \Delta r, 0), r_{\max})$

9 Update replay field: $R'_{u,v} = \frac{\left(H'_{x,y} - H_{x,y}\right)\exp\left[-2\pi i\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right)\right]}{\sqrt{N_xN_y}}$

10 Calculate new error: $E'_{u,v} = \left|T_{u,v} - R'_{u,v}\right|^2$

end

$$\begin{aligned}\Delta E' &= \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\Delta E'_{u,v} \\ &= \Delta r^2 - \frac{2\Delta r}{\sqrt{N_xN_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E_{u,v}}\cos\beta_{u,v}\end{aligned} \tag{5.33}$$

Taking $^{\mathrm{d}\Delta E'}/_{\mathrm{d}\Delta r} = 0$ to find the the value of Δr where $\Delta E'$ is minimum

$$\Delta r = \frac{1}{\sqrt{N_xN_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E_{u,v}}\cos\beta_{u,v} \tag{5.34}$$

The linear nature of this result means that Δr can be capped within the constraints of the SLM. This is shown in Algorithm 5.3.

Performance

Using the target images of Section 5.4.2 with a simulated 2^8 level amplitude SLM gives the performance graph as shown in Figure 5.24. This results in an approximately $2\times$ improvement in convergence speed to reach a given target error.

The algorithm reconstruction is shown in Figure 5.25 showing target image (left), DS (centre) and HPS (right). The SSIM values given are calculated with a dynamic range of 1.0.

Very similar performance improvements were seen when used in place of a simulated annealing algorithm where HPS consistently outperformed SA approximately $2\times$ in terms of iterations required to reach a given target error. While this is interesting, it should be noted that phase sensitive amplitude holography is unlikely to be a common paradigm and this result is presented primarily for completeness.

5.4.6 Amplitude Modulated SLM, Phase Insensitive Replay Field

The final Fraunhofer case to consider is that of a amplitude modulated SLM with a phase insensitive replay field.

Derivation

The phase insensitive amplitude modulated behaviour is similar to the phase insensitive phase modulated case. The problem geometry is shown in Figure 5.16d. Working again in terms of $\Delta r' = r' - r$ where r and r' are the old and new pixel magnitudes respectively.

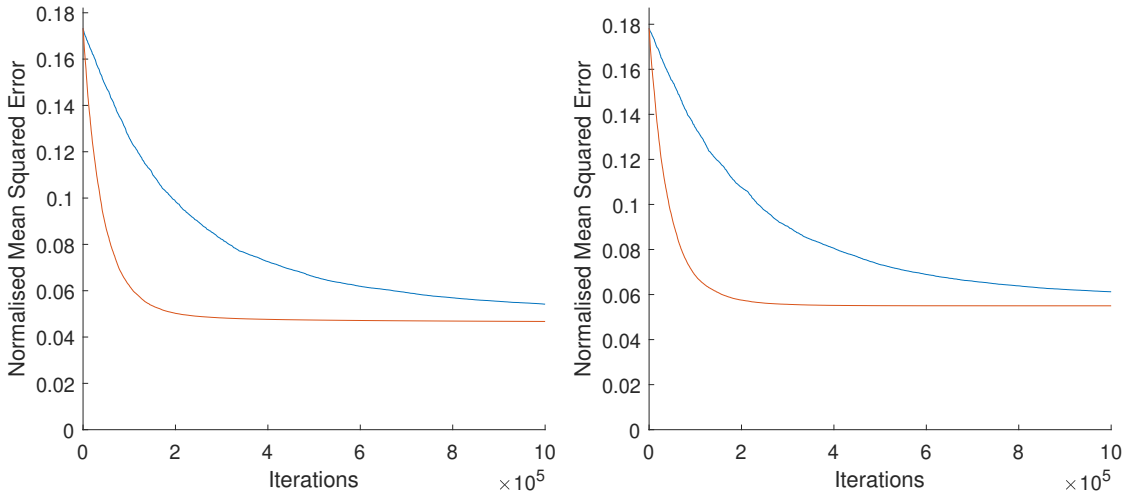


Fig. 5.24 Comparison of Direct Search (blue) against amplitude modulated, phase sensitive HPS (orange).

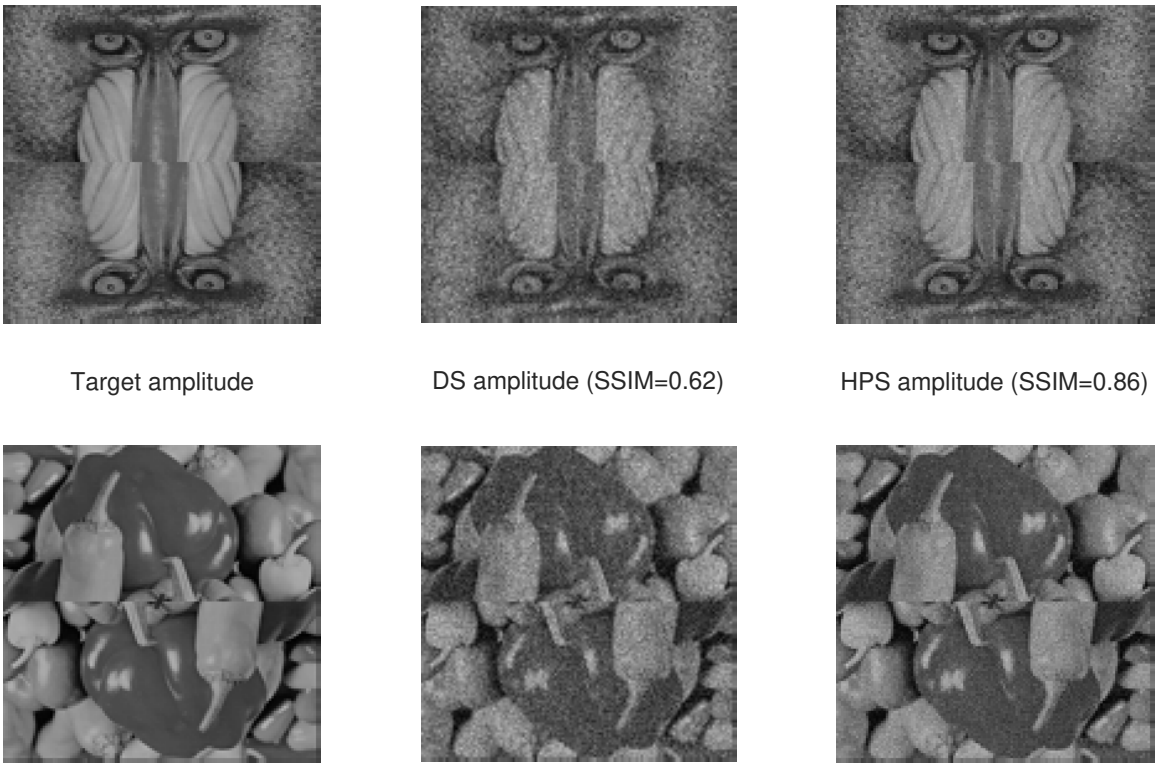


Fig. 5.25 Comparison of Direct Search (centre) against amplitude modulated, phase sensitive HPS (right).

$$
\begin{aligned}
E'_{u,v} &= \left[|T_{u,v}| - \left|R'_{u,v}\right|\right]^2 \\
&= \left[|T_{u,v}| - \sqrt{|R_{u,v}|^2 + \frac{\Delta r^2}{N_x N_y} + 2|R_{u,v}|\frac{\Delta r}{\sqrt{N_x N_y}}\cos\beta}\right]^2 \\
\Rightarrow \Delta E'_{u,v} &= E'_{u,v} - E_{u,v} \\
&= \frac{\Delta r^2}{N_x N_y} + 2|R_{u,v}|\frac{\Delta r}{\sqrt{N_x N_y}}\cos\beta + 2|T_{u,v}||R_{u,v}| \\
&\quad - 2|T_{u,v}||R_{u,v}|\sqrt{|R_{u,v}|^2 + \frac{\Delta r^2}{N_x N_y} + 2|R_{u,v}|\frac{\Delta r}{\sqrt{N_x N_y}}\cos\beta}
\end{aligned} \tag{5.35}
$$

Unlike in Section 5.4.4, there is no easy Taylor substitution. Instead it is assumed that $\Delta r^2 + 2|R_{u,v}|\Delta r\cos\beta$ is smaller than $|R_{u,v}|^2$. This assumption can be seen to be valid for almost all non-zero target replay field values. For the *Mandrill* test image this results in $> 99.99\%$ of pixels being valid. For different amplitude distributions this assumption becomes less valid. Fortunately for us, however, the system is insensitive to such pixels as they have near zero magnitude and this is further improved by the square relationship between intensity and amplitude. In the tests run here, less than 0.02% of pixels gave greater than 1% error in target value due to this assumption.

As a result

$$
\begin{aligned}
\Delta E'_{u,v} &= E'_{u,v} - E_{u,v} \\
&= (1 - |T_{u,v}|)\left(\frac{\Delta r^2}{N_x N_y} + 2|R_{u,v}|\frac{\Delta r}{\sqrt{N_x N_y}}\cos\beta\right)
\end{aligned} \tag{5.36}
$$

where $\beta_{u,v}$ is again given by

$$
\beta_{u,v} = \theta_{u,v} - \angle(T_{u,v} - R_{u,v}) = -2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) - \angle(T_{u,v} - R_{u,v}) \tag{5.37}
$$

Summing

Algorithm 5.4: Amplitude Modulated, Phase Insensitive HPS

Input: Target $T_{u,v}$, current replay $R_{u,v}$ and number of iterations N

Randomise target image phase: $R'_{u,v} = |T_{u,v}| \angle \text{Rand}[0, 2\pi]$

Back-propagate the target to the diffraction plane: $H_{u,v} = \mathscr{F}^{-1}\{R'\}_{u,v}$

Quantise the resultant hologram: $H' = \text{quantise}(H_{u,v})$

Generate initial replay field: $R_{u,v} = \mathscr{F}\{H'\}_{u,v}$

Generate initial error: $E_{u,v} = \left|T_{u,v}^2 - R_{u,v}^2\right|$

for $n \leftarrow 1$ **to** N **do**

  Select a random pixel x, y

  Calculate constant: $\beta_{u,v} = -2\pi\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right) - \angle(T_{u,v} - R_{u,v})$

  Calculate new pixel magnitude: $\Delta r = \frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}(1-T_{u,v})|R_{u,v}|\cos\beta}{\sqrt{N_x N_y}}$

  Calculate new pixel value: $H'_{x,y} = \min(\max(H'_{x,y} + \Delta r, 0), r_{\max})$

  Update replay field: $R'_{u,v} = \frac{\left(H'_{x,y} - H_{x,y}\right)\exp\left[-2\pi i\left(\frac{ux}{N_x} + \frac{vy}{N_y}\right)\right]}{\sqrt{N_x N_y}}$

  Calculate new error: $E'_{u,v} = \left(|T_{u,v}| - \left|R'_{u,v}\right|\right)^2$

end

$$\begin{aligned}\Delta E' &= \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\Delta E'_{u,v} \\ &= (1 - \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}T_{u,v})\Delta r^2 + \frac{2\Delta r}{\sqrt{N_x N_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}(1-T_{u,v})|R_{u,v}|\cos\beta \end{aligned} \tag{5.38}$$

Taking ${}^{\mathrm{d}\Delta E'}/_{\mathrm{d}\Delta r} = 0$ to find the the value of Δr where $\Delta E'$ is minimum

$$\Delta r = \frac{1}{\sqrt{N_x N_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}(1-T_{u,v})|R_{u,v}|\cos\beta \tag{5.39}$$

The linear nature of this result means that Δr can be capped within the constraints of the SLM. This translates into Algorithm 5.4

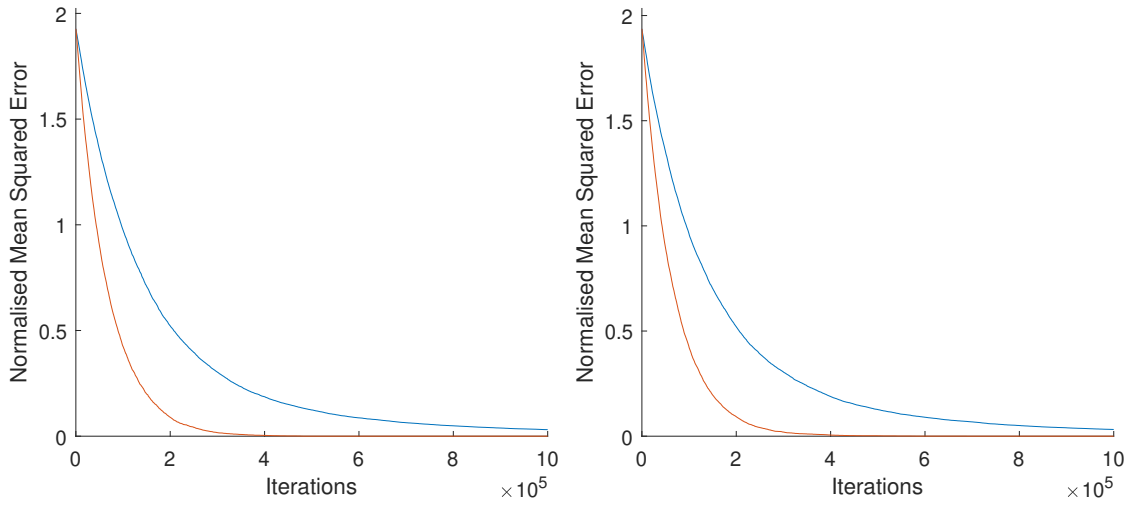


Fig. 5.26 Comparison of Direct Search (blue) against amplitude modulated, phase insensitive HPS (orange).

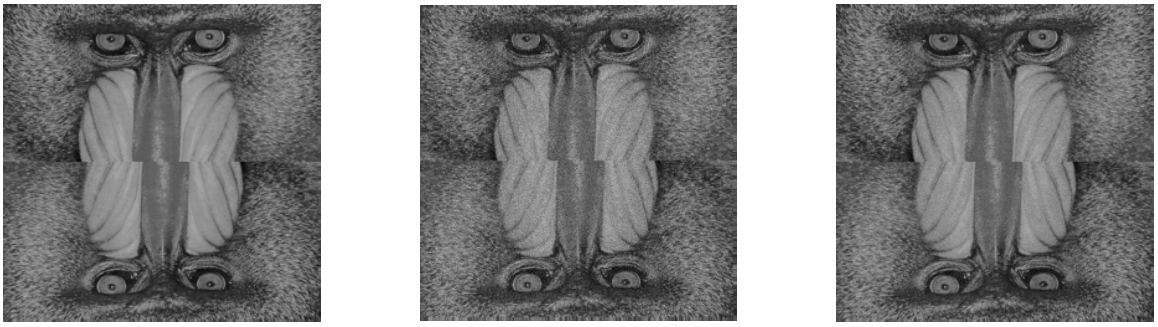


Fig. 5.27 Comparison of Direct Search (centre) against amplitude modulated, phase insensitive HPS (right).

Performance

Tests of convergence for a 256×256 *Mandrill* test image on a 2^8 level amplitude SLM gives the performance graph as shown in Figure 5.26. This again results in an approximately $2\times$ improvement in convergence iterations to reach a given target error.

The algorithm reconstruction is shown in Figure 5.27 showing target image (left), DS (centre) and HPS (right). The SSIM values given are calculated with a dynamic range of 1.0.

5.4.7 Fresnel Domain

Fortunately, the Fresnel transform variants of HPS turn out to have a very similar form to their Fraunhofer counterparts. The only distinction is the addition of the quadratic phase term to $C_{u,v}$ and $\beta_{u,v}$

$$C_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)+\angle R^{\dagger}_{u,v}+e^{\frac{i\pi}{\lambda z}(x^2+y^2)}$$
$$\beta_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)-\angle(T_{u,v}-R_{u,v})+e^{\frac{i\pi}{\lambda z}(x^2+y^2)} \tag{5.40}$$

where z is the perpendicular separation between diffraction field and replay field and λ is the illumination wavelength.

5.4.8 Discussion

The required relationships for the different HPS variants are summarised in Table 5.1 along with algorithm flowcharts in Figure 5.28. There are a significant number of points that should be discussed.

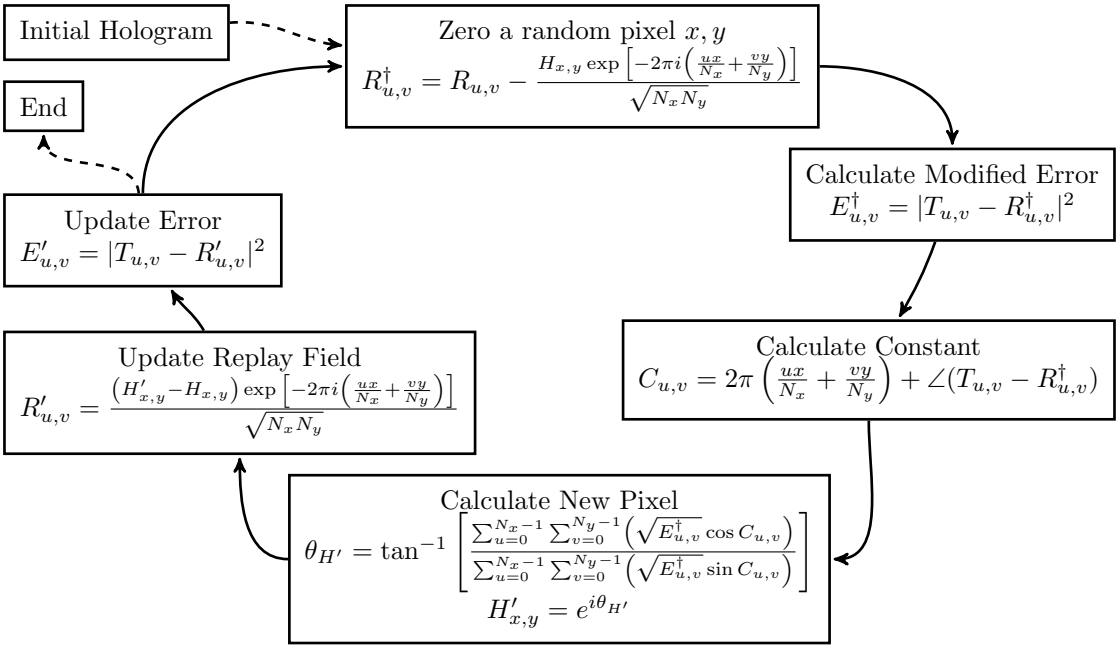
Firstly, competitor algorithm families should be considered. HSAs such as DS can be used to generate some of the best quality holograms, albeit at the expense of slower generation times and HPS uniformly out-performed DS for every case discussed in this paper in both speed and convergent quality.

Iterative algorithms - for example Gerchberg-Saxton[50] - are available for many relatively smooth systems and should be expected to be significantly faster than HSAs including HPS. This comes at the expense of final image quality where HPS is expected to give the best performance. For more-discontinuous systems with lower numbers of modulation levels, iterative algorithms can fail to converge and HSAs become a suitable alternative. In this case HPS may be expected to offer better performance in speed as well as quality.

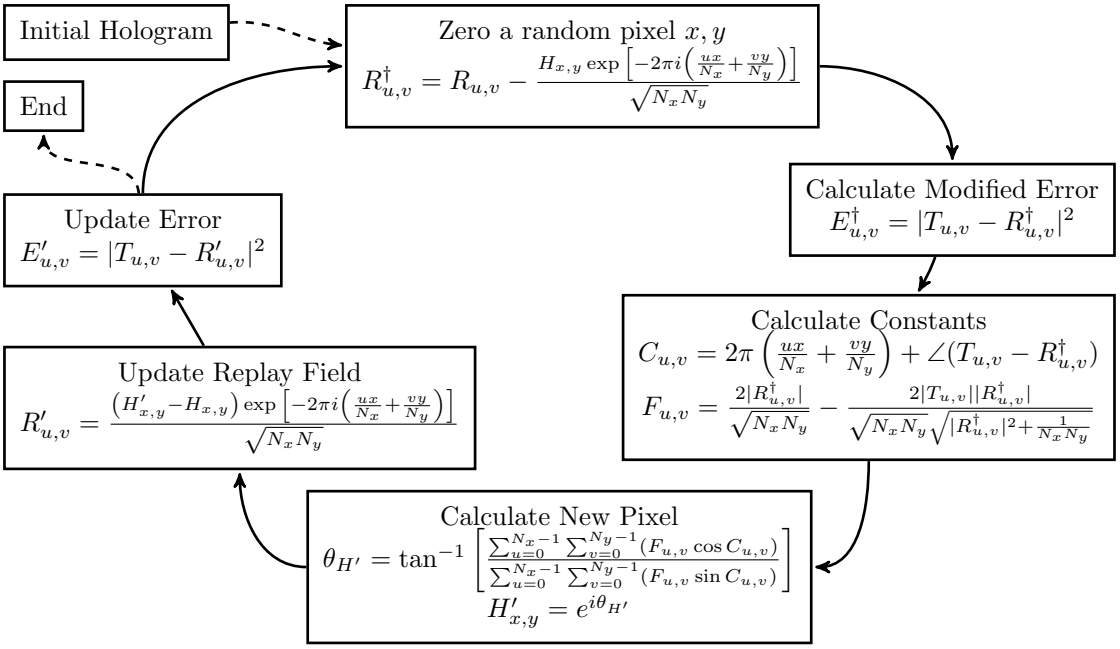
Secondly, the computational performance of HPS should be considered as HPS required approximately $70-80\%$ more time per iteration though this dropped to as little as 10% in the case of larger images where computation was memory bound. This increased iteration time should be taken into account when selecting an appropriate algorithm.

Table 5.1 Mathematical relationships for different HPS variants

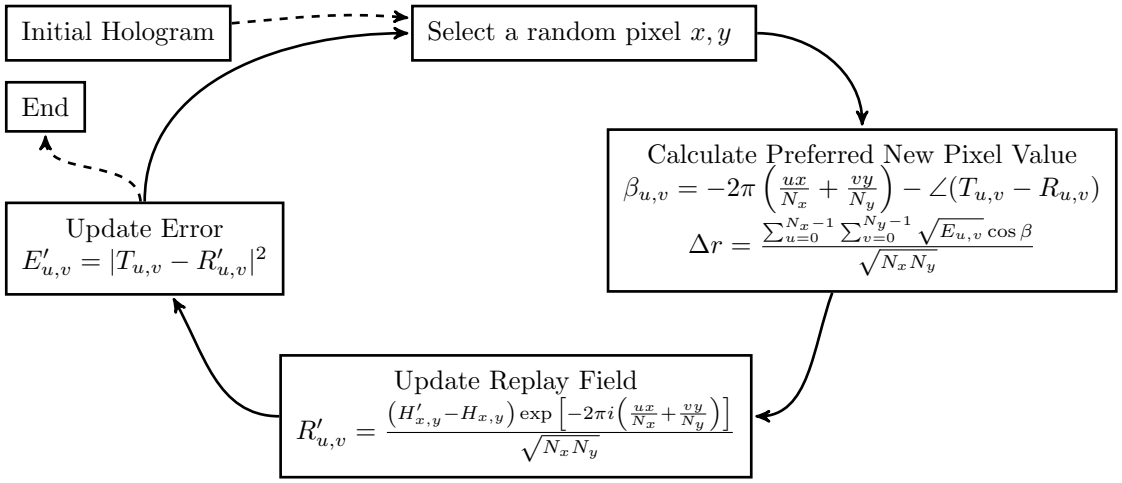
Phase Modulated SLM, Phase Sensitive Replay Field, Fraunhofer Domain	Phase Modulated SLM, Phase Sensitive Replay Field, Fresnel Domain
$C_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)+\angle R^{\dagger}_{u,v}$	$C_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)+\angle R^{\dagger}_{u,v}+e^{\frac{i\pi}{\lambda z}(x^2+y^2)}$
$\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\cos C_{u,v}\right)}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\sin C_{u,v}\right)}\right]$	$\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\cos C_{u,v}\right)}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\sqrt{E^{\dagger}_{u,v}}\sin C_{u,v}\right)}\right]$
Phase Modulated SLM, Phase Insensitive Replay Field, Fraunhofer Domain	Phase Modulated SLM, Phase Insensitive Replay Field, Fresnel Domain
$C_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)+\angle R^{\dagger}_{u,v}$	$C_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)+\angle R^{\dagger}_{u,v}+e^{\frac{i\pi}{\lambda z}(x^2+y^2)}$
$F_{u,v} = \frac{2\left\|R^{\dagger}_{u,v}\right\|}{\sqrt{N_xN_y}}-\frac{2\left\|T_{u,v}\right\|\left\|R^{\dagger}_{u,v}\right\|}{\sqrt{N_xN_y}\sqrt{\left\|R^{\dagger}_{u,v}\right\|^2+\frac{1}{N_xN_y}}}$	$F_{u,v} = \frac{2\left\|R^{\dagger}_{u,v}\right\|}{\sqrt{N_xN_y}}-\frac{2\left\|T_{u,v}\right\|\left\|R^{\dagger}_{u,v}\right\|}{\sqrt{N_xN_y}\sqrt{\left\|R^{\dagger}_{u,v}\right\|^2+\frac{1}{N_xN_y}}}$
$\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(F_{u,v}\cos C_{u,v}\right)}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(F_{u,v}\sin C_{u,v}\right)}\right]$	$\theta_{H'} = \tan^{-1}\left[\frac{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(F_{u,v}\cos C_{u,v}\right)}{\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(F_{u,v}\sin C_{u,v}\right)}\right]$
Amplitude Modulated SLM, Phase Sensitive Replay Field, Fraunhofer Domain	Amplitude Modulated SLM, Phase Sensitive Replay Field, Fresnel Domain
$\beta_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)-\angle(T_{u,v}-R_{u,v})$	$\beta_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)-\angle(T_{u,v}-R_{u,v})+e^{\frac{i\pi}{\lambda z}(x^2+y^2)}$
$\Delta r = \frac{1}{\sqrt{N_xN_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E_{u,v}}\cos\beta$	$\Delta r = \frac{1}{\sqrt{N_xN_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\sqrt{E_{u,v}}\cos\beta$
Amplitude Modulated SLM, Phase Insensitive Replay Field, Fraunhoferr Domain	Amplitude Modulated SLM, Phase Insensitive Replay Field, Fresnel Domain
$\beta_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)-\angle(T_{u,v}-R_{u,v})$	$\beta_{u,v} = 2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)-\angle(T_{u,v}-R_{u,v})+e^{\frac{i\pi}{\lambda z}(x^2+y^2)}$
$\Delta r = \frac{1}{\sqrt{N_xN_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}(1-T_{u,v})\lvert R_{u,v}\rvert\cos\beta$	$\Delta r = \frac{1}{\sqrt{N_xN_y}}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}(1-T_{u,v})\lvert R_{u,v}\rvert\cos\beta$



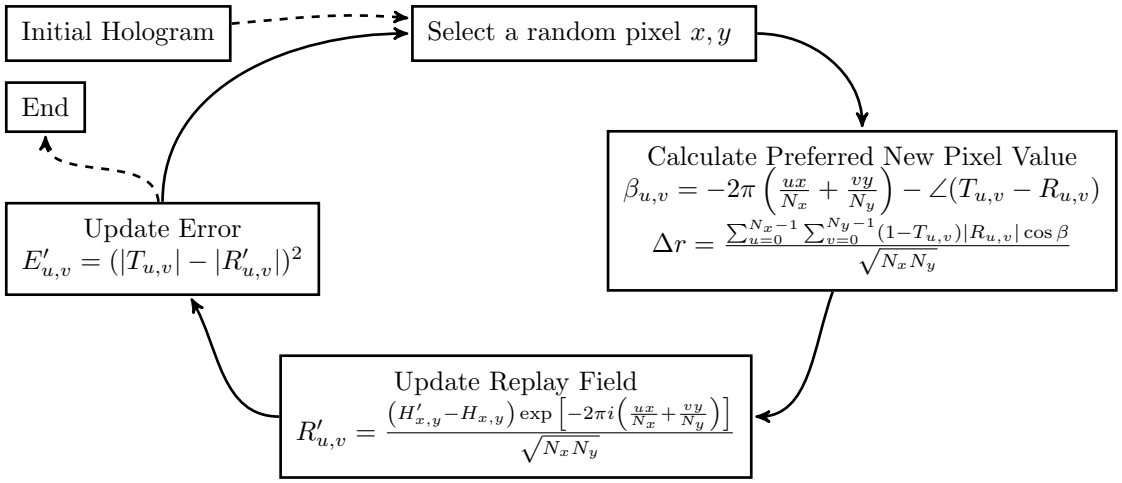
(a) Phase modulated, phase sensitive



(b) Phase modulated, phase insensitive



(c) Amplitude modulated, phase sensitive



(d) Amplitude modulated, phase insensitive

Fig. 5.28 HPS algorithms for phase and amplitude modulated SLMs in the case of phase sensitive or phase insensitive replay field

Thirdly, this study has been purely mathematical in nature. Account has not been taken for real-world imperfections such as lens aberration, non-flatness or speckle. The author suggest that these effects are likely to affect all HSAs similarly but recommend further study of the sensitivity of different algorithms to real-world errors.

Fourthly, the image quality metric used here is MSE. SSIM has seen increased use in recent years as it more closely corresponds to the human eye behaviour. While recent authors have argued for a closer relationship between MSE and SSIM than thought previously [287] it is acknowledged that this is a weakness in the HPS method and further investigation is required.

Finally, regarding the complexity of the HPS method. While in certain situations HPS offers significant performance improvements over rival techniques, the increased complexity and reduced generality will require a greater level of expertise than alternative techniques.

5.4.9 Summary

This work has presented seven new variants on the holographic predictive search (HPS) algorithm which are summarised in Table 5.1. By using this, prescient search techniques can be used for a wide range of optical systems in both the far- and mid-field. Different modulation schemes and replay field constraints have all been discussed.

When compared to direct search and simulated annealing algorithms, HPS has been shown to be over $10\times$ faster than its competitors in specific cases at the expense of increased complexity and reduced flexibility. HPS also offers the best convergent error quality. Variants on the HPS algorithm have been presented for a range of optical configurations and the relative advantages and disadvantages presented.

5.5 Sympathetic Quantisation[4]

5.5.1 Introduction

This section sets out a new concept in CGH generation termed sympathetic quantisation (SQ). SQ is a novel approach that exploits the underlying relationships of Fraunhofer and Fresnel diffraction in order to improve quantisation behaviour in CGH. A single example of this approach is developed called soft sympathetic quantisation (SSQ) which is designed for use with single-iteration time-multiplexed algorithms. With the addition of SSQ to OSPR or STTM - Section 2.4.5 and Section 5.3 - significant quality improvements can be achieved at very low additional cost on generation time.

The driving limitation of CGH generation is the modulation step where the aperture function is constrained by the limited modulation capabilities of the SLMs used [268]. For example, phase-modulating SLMs can only vary the phase of a pixel with the amplitude remaining unchanged. Often the modulation constraint is worsened by the digital nature of SLMs where the continuously modulated hologram is quantised to the discrete energy levels achievable by the device. For example, 8-bit phase only SLMs are constrained to quantisation angle steps of $2\pi/256$ radians on the Argand circle.

The modulation and quantisation scheme used in these algorithms is nearest neighbour quantisation (NNQ). Here the ideal hologram value is changed to the closest achievable state in $\mathbb{C}$. If the pixels in the hologram can be assumed independent of each other then it can be shown that this is statistically the best procedure [266, 267]. This section presents an alternative approach that is aware of the correlated relationship between individual SLM pixels and uses this to improve the quality of single-iteration holograms.

5.5.2 Sympathetic Quantisation Approach

The update step discussed in Section 2.4.4, (2.72)

[4]The contents of this section were previously included as part of the following publication:

> Peter J. Christopher, Ralf Mouthaan and Timothy D. Wilkinson, “Sympathetic quantisation - a new approach to hologram quantisation", *Optics Communications*, (2020), Volume 473, DOI: 10.1016/j.optcom.2020.125883

$$\Delta R_{u,v} = \frac{1}{\sqrt{N_x N_y}} \Delta H_{x,y} e^{\left[-2\pi i\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)\right]} \tag{5.41}$$

iis the fact that modifying defined pairs of pixels synchronously can allow us limited control over the error caused by modulation or quantisation. For example if two pixels - $H_{x,y}$ and $H_{x\pm\frac{N_x}{2},y}$ - are taken with identical y coordinates and with x coordinates separated by $N_x/2$ 100% of the error can be localised to 50% of the columns with 50% remaining error free. Analytically this can be seen from

$$\Delta R_{u,v} = \frac{1}{\sqrt{N_x N_y}} \left[\Delta H_{x,y} e^{\left[-2\pi i\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)\right]} + \Delta H_{x\pm\frac{N_x}{2},y} e^{\left[-2\pi i\left(\frac{u(x\pm\frac{N_x}{2})}{N_x}+\frac{vy}{N_y}\right)\right]} \right] \tag{5.42}$$

which can be seen to cancel for values of $u = \pm 2\pi$ provided $\Delta H_{x,y} = \Delta H_{x\pm\frac{N_x}{2},y}$. This result is of little practical use but, as shall be shown later, depending on the location relationship between the pixels, the principle of sympathetic pixel quantisation can be exploited in a number of interesting ways.

Two common features of hologram applications are relevant for this exploitation. Firstly, human vision is phase insensitive with the eye seeing the intensity of the light given by $|R_{u,v}|^2$. This allows judicious phase control to shift error into phase terms where the visual quality is not effected. The second application is spatial. Many applications are only interested in portions of the replay field, allowing error to be moved to the portions of the replay field of lower concern.

The greatest challenge to utilising SQ is its mathematical complexity. Unless care is taken, expressions for paired movements become quartic and therefore computationally expensive. In the remainder of this work a single example of SQ is developed which uses judicious formulation to avoid quartic solutions and illustrates the power of the SQ approach. It will be shown that, for real-time applications, this allows significant improvements on single-frame algorithms such as OSPR and STTM [43, 267].

5.5.3 Soft Sympathetic Quantisation

In order to demonstrate SQ in action, soft sympathetic quantisation (SSQ) will be discussed. SSQ is applicable to phase modulated, phase insensitive hologram generation where only the

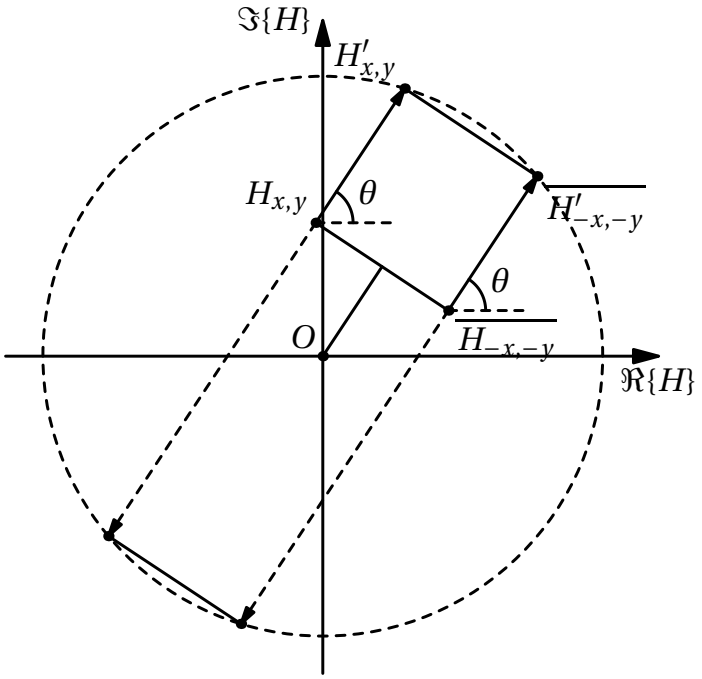


Fig. 5.29 Soft sympathetic quantisation for continuous phase devices

intensity of the replay is of concern and the replay phase is insignificant. This is commonly found in display applications due to the phase insensitivity of the eye. If pairs of pixels are adjusted, $H_{x,y}$ and $H_{-x,-y}$, at locations rotationally symmetric around the origin

$$\Delta R_{u,v} = \frac{1}{\sqrt{N_x N_y}} \left[\Delta H_{x,y} e^{\left[-2\pi i \left(\frac{ux}{N_x} + \frac{vy}{N_y}\right)\right]} + \Delta H_{-x,-y} e^{\left[-2\pi i \left(\frac{-ux}{N_x} + \frac{-vy}{N_y}\right)\right]} \right] \tag{5.43}$$

It can be seen that $\Delta R_{u,v}$ necessarily has angle 0 or π independently of the value of u, v, x or y provided $\Delta H_{x,y} = \overline{\Delta H_{-x,-y}}$ for all u, v. Provided these conditions are kept, the values of $\Delta H_{x,y}$ and $\Delta H_{-x,-y}$ can be adjusted freely.

In simple terms, if a pair of pixels are taken with positions symmetric around the origin and modified so that the change in one pixel is the complex conjugate of the change of the other replay field changes due to quantisation can be localised to lie on a single line on the Argand diagram.

Mathematical Formulation

The standard form of quantisation, NNQ, is given as

$$H'_{x,y} = \exp(2\pi i \angle H_{x,y}), \quad H'_{-x,-y} = \exp(2\pi i \angle H_{-x,-y}) \tag{5.44}$$

Note that here $\angle$ represents the phase operator and $a\angle b$ represents rotating a by b radians.

To meet the SSQ constraints, this is replaced this with a relationship for new pixel values $H'_{x,y}$ and $H'_{-x,-y}$ where

$$H'_{x,y} - H_{x,y} = \overline{H'_{-x,-y}} - \overline{H_{-x,-y}}, \quad \left|H'_{x,y}\right| = \left|H'_{-x,-y}\right| = 1 \tag{5.45}$$

This can be represented geometrically as shown in Figure 5.29 where the modulation problem becomes the one of transforming the chord between $H_{x,y}$ and $H_{-x,-y}$ in order to lie on the circle when $\left|H_{x,y} - H_{-x,-y}\right| \leq 2/\sqrt{N_x N_y}$ or to lie through the origin in the case $\left|H_{x,y} - H_{-x,-y}\right| > 2/\sqrt{N_x N_y}$.[5]

It can be shown using the intersecting chords theorem that vectors c and m can be chosen so that

$$c = \frac{H_{x,y} + \overline{H_{-x,-y}}}{2}$$

$$m = \pm i(H_{x,y} - \overline{H_{-x,-y}})\sqrt{\frac{1}{\left|H_{x,y} - \overline{H_{-x,-y}}\right|^2} - \frac{1}{4}} \tag{5.46}$$

to give

$$H'_{x,y} = H_{x,y} + m - c, \quad H'_{-x,-y} = H_{-x,-y} + \overline{m} - c \tag{5.47}$$

This approach can be executed in parallel and is negligible in execution time when compared to the FFT element.

Note also, that this formulation ceases to work for $\left|H_{x,y} - \overline{H_{-x,-y}}\right| > 2/\sqrt{N_x N_y}$ as there is no longer a way of moving the pixel pairs to the circle while still satisfying the constraints. In these cases the points are adjusted so that the point equidistant between them lie at the origin. For the test images used this occurred less than 0.1% of the time.

[5]The author would like to thank Dr Colin Christopher for initial conversations regarding geometric interpretation of the SQ approach.

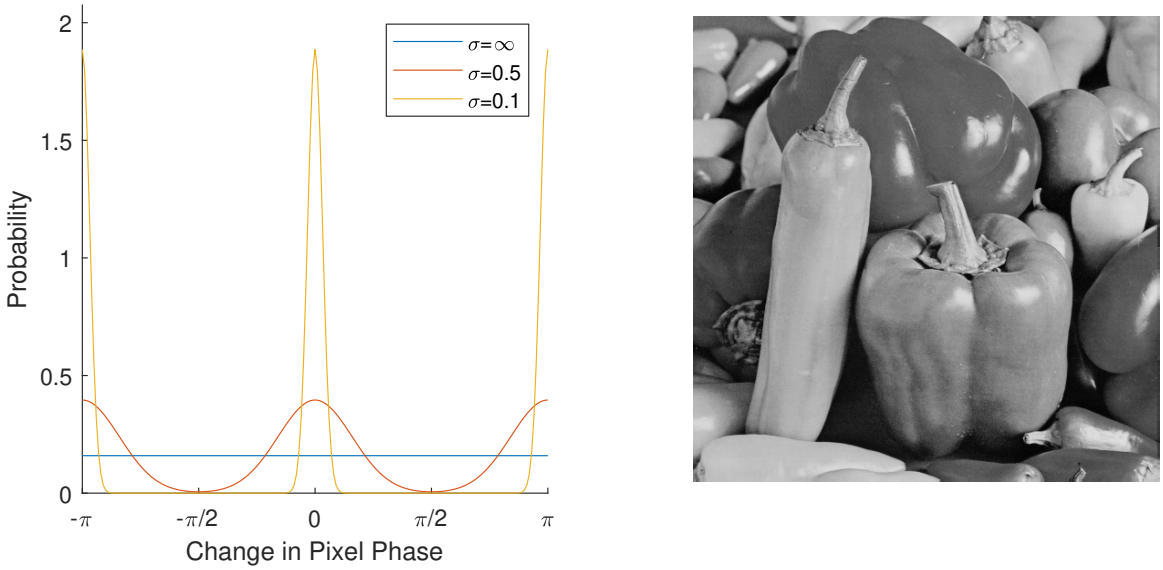


Fig. 5.30 Probability densities for phase randomisation (left) and *Peppers* test image (right).

Phase randomisation

Before discussing performance a digression must be made to talk about *phase randomisation*. Typically for phase insensitive holograms, the seed image used for the algorithm has a uniformly randomised phase profile. This reduces edge enhancement and serves to smooth the spectral profile. For algorithms like Gerchberg-Saxton (GS) [50], this is only significant for the first few iterations and makes little to no difference to convergent behaviour. For single-iteration algorithms like OSPR and STTM phase randomisation is more important as there is no iterative convergence process.

For SSQ to work, however, the seed phases are required to lie near to or on a single axis on the Argand diagram. In doing so it is ensured that any errors introduced during quantisation lie perpendicular to this axis, i.e. along the azimuthal direction of the Argand diagram, corresponding to phase errors to which the eye is not sensitive. Instead of the more traditional uniformly distributed seed phase, a narrow band phase randomisation approach is used as shown in Figure 5.30 (left) which shows the dual von Mises distribution used.

Behaviour

In order to better visually understand the change in pixels, a selection 10 random pairs of hologram pixels are taken as shown in Figure 5.31 (right). The relationships in (5.46)-(5.47)

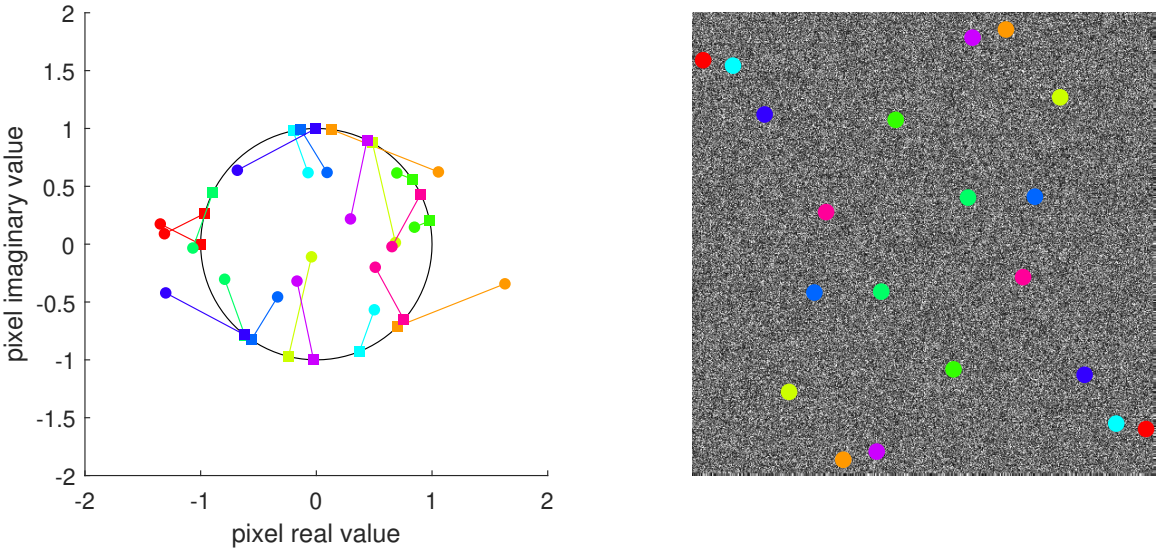


Fig. 5.31 Change of 10 randomly selected pixel pairs (left) for the *Peppers* test image showing the movement of the pixel values with starting points shown as circles and end points as squares. The pixel locations on the generated hologram are shown right.

are used to quantise the pixels. Figure 5.31 (left) showing the starting values as circles and the final values as squares.

Algorithm

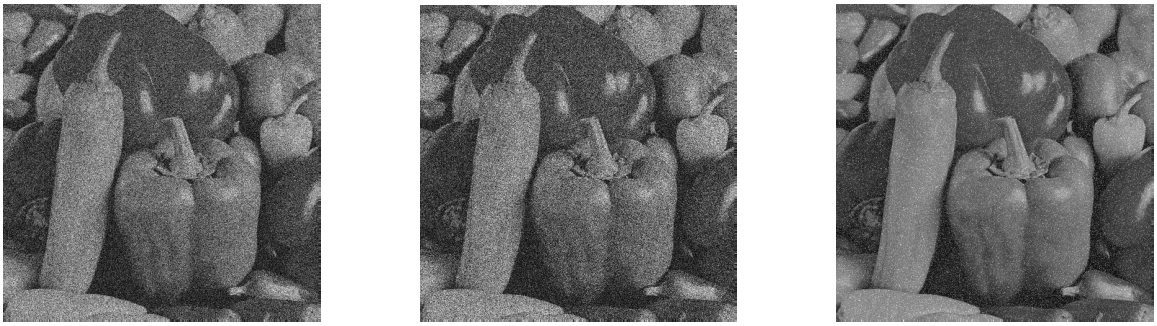
Single Iteration

Figure 5.32 shows holograms produced using full phase randomisation with NNQ (left), narrow band ($\sigma = 0.05$) randomisation with NNQ (centre) and narrow band ($\sigma = 0.05$) randomisation with SSQ (right).

The mean squared error (MSE) and structural similarity index (SSIM) performance metrics are calculated according to Section 2.4.1.

Case (b) is unlikely to be used in a real-world system but is included to highlight the competing factors. Moving from full randomisation to narrow band randomisation accounts for the decrease in quality between (a) and (b). This is more than compensated for by the addition of SSQ between (b) and (c). Similar results are seen when the algorithm is applied to other standard test images such as *Peppers* and *Camera Man*.

Multiple Iterations



NMSE=0.103, SSIM=0.0928 NMSE=0.125, SSIM=0.0801 NMSE=0.0491, SSIM=0.151

Fig. 5.32 Comparison of the initial inverse transform with full phase randomisation with NNQ (left) narrow band ($\sigma = 0.05$) randomisation with NNQ (centre) and narrow band ($\sigma = 0.05$) randomisation with SSQ (right). The SSIM measurements assume a dynamic range equal to 1. The SLM is assumed to have 256 levels.

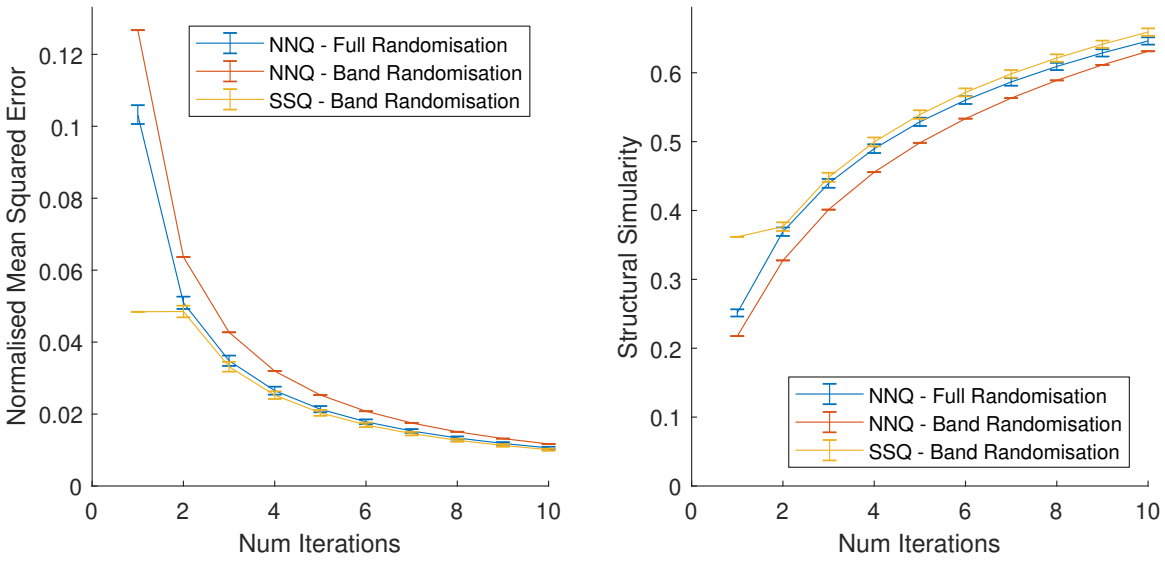


Fig. 5.33 Convergence of the GS algorithm using SSQ compared to NNQ with full and narrow band ($\sigma = 0.05$) randomisation. MSE is shown left with SSIM right. The SSIM measurements assume a dynamic range equal to 1. The SLM is assumed to have 256 levels.

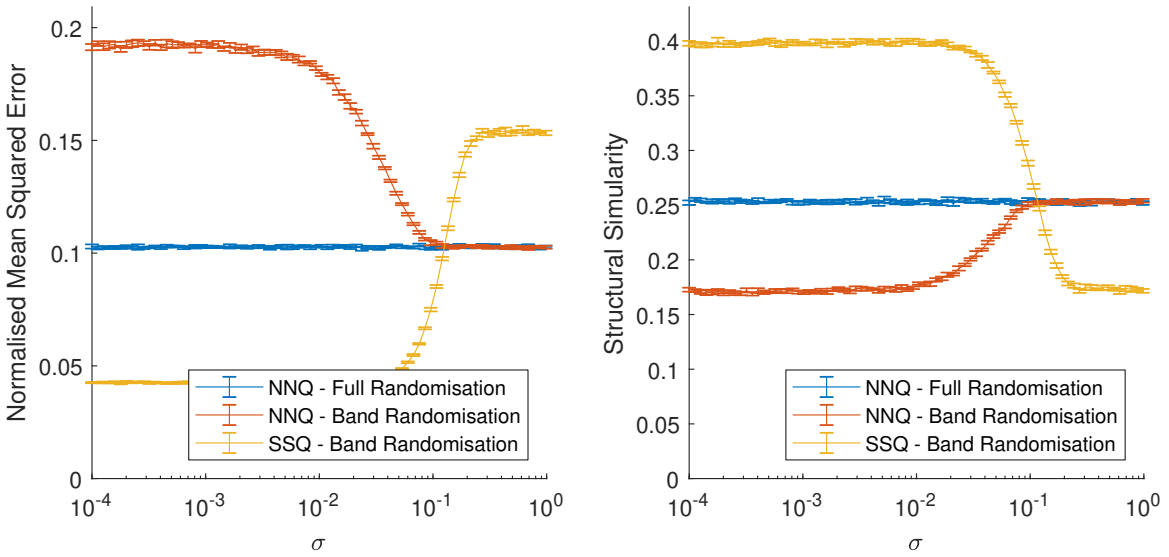


Fig. 5.34 Comparison of SSQ performance compared to NNQ with full and banded randomisation against band width. MSE is shown left with SSIM right. The SSIM measurements assume a dynamic range equal to 1. The SLM is assumed to have 256 levels. Values are taken as being the mean of 20 independent runs with error bars showing two standard deviations.

SSQ fails to continue working when applied to iterative algorithms such as GS. As shown in Figure 5.33, the first iteration offers significant performance benefits but the advantages disappear after the first iteration. The reason for this is that only the first iteration is done with a narrow band randomised phase distribution. Once phase distribution is more varied, SSQ becomes detrimental and convergence is worse than the NNQ case.

This suggests that SSQ is applicable primarily to single iteration approaches such as OSPR and STTM where individual hologram quality is sacrificed in favour of faster generation speed for real-time displays. Here SSQ shows strong improvements in both MSE and SSIM.

Choice of σ

This prompts the question, what value of sigma should be chosen? Too high and the initial replay phase is no longer sufficiently uniform, too low and edge enhancement effects may well begin to dominate. Figure 5.34 suggests that σ can be reduced to near zero.

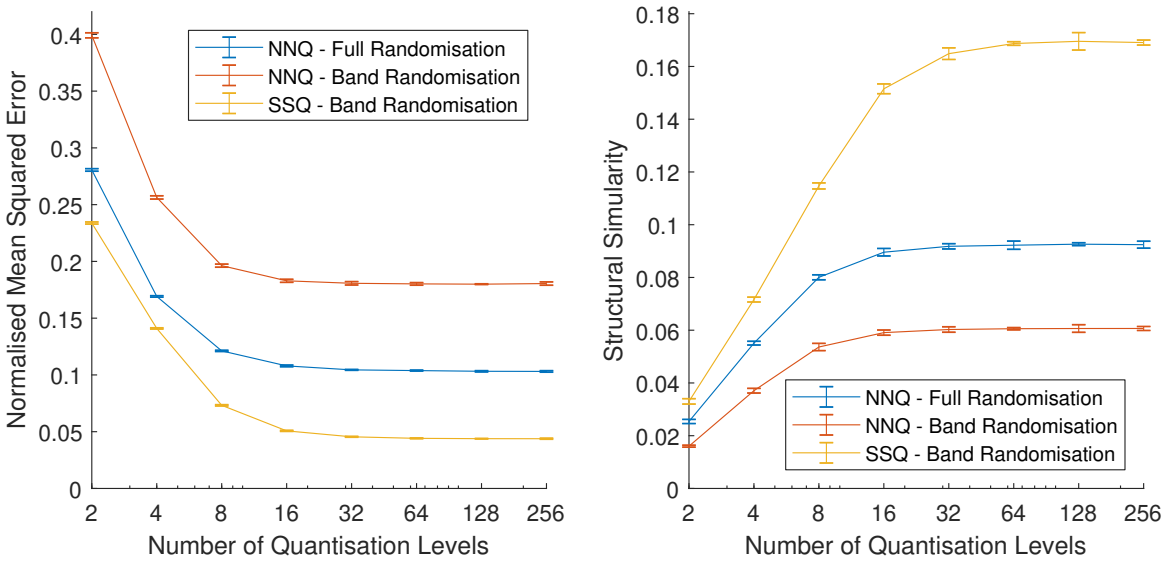


Fig. 5.35 Comparison of SSQ performance compared to NNQ with full and banded randomisation against number of quantisation levels. MSE is shown left with SSIM right. The SSIM measurements assume a dynamic range equal to 1. The narrow band randomisation is taken with $\sigma = 0.01$. Values are taken as being the mean of 20 independent runs with error bars showing two standard deviations.

Fresnel Diffraction

The discussion so far has focussed on Fourier or Fraunhofer holograms. Fresnel holograms can be represented in a similar manner with the addition of a quadratic phase term

$$R_{u,v} = \underset{\text{Fresnel}}{\mathscr{F}}\{H_{x,y}\} = \underset{\text{Fraunhofer}}{\mathscr{F}}\{H_{x,y}e^{\frac{i\pi}{\lambda z}(x^2+y^2)}\} \tag{5.48}$$

where λ is the illumination wavelength. Fortunately the rotational symmetry of SSQ means that (5.43) still applies and by extension (5.46)-(5.47).

Quantisation Levels

All the results presented so far have been for the case of SLMs with 256 quantisation levels and it is worth investigating the case with lower numbers of quantisation levels. Figure 5.35 shows a comparison of MSE and SSIM against number of quantisation levels for the first iteration.

Fig. 5.36 Experimental implementation of Figure 5.32 with full phase randomisation with NNQ (left) narrow band ($\sigma = 0.05$) randomisation with NNQ (centre) and narrow band ($\sigma = 0.05$) randomisation with SSQ (right). Test image used is taken from the USC-SIPI database[49] and is shown in Figure 2.12. The right hand image can be seen to present exhibit lower noise and improved contrast when compared to the left hand image. The centre shows the effect of just reducing phase randomisation without the use of SSQ techniques.

This shows that SSQ offers the greatest performance improvements for higher numbers of quantisation levels but still offers performance benefits for low numbers of quantisation levels.

Experimental

Figure 5.36 shows the same holograms shown in Figure 5.32 displayed using the experimental system described in Section 5.1.1 with only 512×512 pixels of the Jasper SLM being enabled to ensure a fair comparison.

Similar to Section 5.5.3, case (b) is unlikely to be used in a real-world system but is included to highlight the effect of removing phase randomisation without the use of technique such as SSQ. Moving from full randomisation to narrow band randomisation accounts for the decrease in quality between (a) and (b). This is more than compensated for by the addition of SSQ between (b) and (c). Similar results are seen when the algorithm is applied to other standard test images such as *Peppers* and *Camera Man*.

While this shows a similar change in visual quality between the individual frames it must be noted that the dominant sources of image degradation appear are not primarily due to the hologram displayed.

Applications and Limitations

For single frame approaches such as OSPR and STTM, SSQ offers the potential to significantly improve both MSE and SSIM. For the example images given, MSE was reduced to under 50% of more traditional approaches while SSIM saw a greater than 50% improvement. This is, unfortunately, limited to only the first iteration of the algorithm.

A number of time-multiplexed algorithms, such as OSPR or STTM only operate in single frame contexts, time averaging many low quality frames. Here speed of generation is paramount and here SSQ offers significant performance benefits.

The computational overhead of SSQ is low with our implementation spending more than 98% of runtime on FFT calculation and SSQ requiring less than 0.2% additional computational overhead. Mathematical complexity is also straight forward with (5.46)-(5.47) only requiring simple algebraic manipulation.

5.5.4 Summary

This section has set out to do two things. Firstly to introduce an alternative approach to hologram quantisation and secondly to present a simple example of this in action.

This work has presented an approach for hologram quantisation called sympathetic quantisation. SQ uses the mathematical formulation of the Fourier transform to adjust pairs of pixel simultaneously during hologram quantisation. This paired movement allows for greater control of the resultant error in the replay field and by extension image quality. By using geometric approaches is it possible avoid the quartic relationships that similar problems often degenerate to.

Significant work is still required to explore alternative formulations in hologram generation. The ability to control the location of replay field error is an exciting opportunity for hologram designers and it is anticipated that this will prove profitable for future study. For example, in fibre mode generation [223] both the amplitude and phase of the replay field are controlled but only for a small central portion. It is anticipated that SQ would allow for a quantisation technique that localised quantisation error to regions outside of the region of interest.

A single example of SQ, soft sympathetic quantisation, has been presented which uses a simple relationship in (5.46)-(5.47) to update pairs of hologram pixels located symmetrically

around the origin in a manner that moves replay error into phase rather than intensity. For the example images given MSE was reduced to under 50% when compared to traditional NNQ while SSIM saw a greater than 50% improvement. For time-multiplexing single-iteration algorithms such as OSPR and STTM this is a significant performance benefit at negligible cost to performance.

Many questions remain worthy of exploration for SSQ. Firstly, combining SSQ with algorithms more advanced than OSPR or STTM is likely to be beneficial. Secondly, understanding the effect of target image magnitude spectrum on performance is expected to be worthwhile. Perhaps the biggest unanswered question is whether this approach can be extended to greater numbers of pixels. It is anticipated that manipulation of 3 or more pixels may allow for further advanced replay noise control opportunities.

5.6 Linear Time[6]

5.6.1 Introduction

A widely used family of algorithms for phase sensitive replay fields are the holographic search algorithms (HSAs) of which the most famous are perhaps DS and SA discussed in Sections 2.4.4 and 11 and summarised in Figure 2.14e and 2.14f. In this section a novel approach to phase sensitive CGH is demonstrated which has been termed *linear time* (LT).

5.6.2 Basic Premise

For the initial investigation it shall be shown that by using known properties of the Fourier transform it is possible to significantly reduce the computation required for generating phase sensitive holograms. Note that this only considers Fraunhofer holograms without a region of interest (ROI), i.e. the entire replay field is to be optimised. This analysis shall be extended to Fresnel holograms and and the inclusion of an RoI later in this section.

[6]The contents of this section were previously included as part of the following publication:

> Peter J. Christopher, Ralf Mouthaan, Miguel El Guendy and Timothy D. Wilkinson, “Linear time algorithm for phase sensitive holography", *Optical Engineering*, (2020), Volume 59, Issue 8, DOI: 10.1117/1.OE.59.8.085104

The Fourier transform operation obeys Parseval's theorem - Section 2.1.5, Eq. 2.37 which corresponds to energy conservation between the diffraction and replay field planes which is the cause of the the $1/\sqrt{N_x N_y}$ terms in Eq. 2.5.

These definitions can be used with Parseval's theorem to obtain a new expression for the MSE metric - Section 2.4.1, (2.68).

$$\begin{aligned} E_{MSE} &= \frac{1}{N_x N_y} \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} |T_{u,v} - R_{u,v}|^2 \\ &= \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} (T_{u,v} - R_{u,v})(\overline{T_{u,v}} - \overline{R_{u,v}}) \\ &= \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} (G_{x,y} - H_{x,y})(\overline{G_{x,y}} - \overline{H_{x,y}}) \\ &= \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \left|G_{x,y} - H_{x,y}\right|^2 \end{aligned} \tag{5.49}$$

The key innovation here is the observation that this allows the value of E_{MSE} to be determined on the SLM side of the transform from $G_{x,y}$ and $H_{x,y}$, and that this avoids the need for repeatedly projecting changes to the replay fields side to calculate the MSE. If the original MSE is known, then the effect of any change can be determined in $O(1)$ rather than the $O(N_x N_y)$ time required for a calculation on the replay field side.

Result: Mean squared error calculation for any phase sensitive Fraunhofer hologram can be done in the diffraction plane.

5.6.3 Linear-Time Holographic Search Algorithm

Crucially, the calculation of $G_{x,y}$ needs to only be done once - before the hologram calculation commences - in other words, there is no longer a need for repeated Fourier transform evaluations at each iteration. While it may appear obvious that (refparseval) necessitates that (2.68) and (5.49) are equivalent, this result has apparently never been used previously for hologram generation. Importantly, if the MSE $E_{MSE}(G,H)$ is known, changing a single pixel in H at coordinates x, y allows the new error to be expressed as

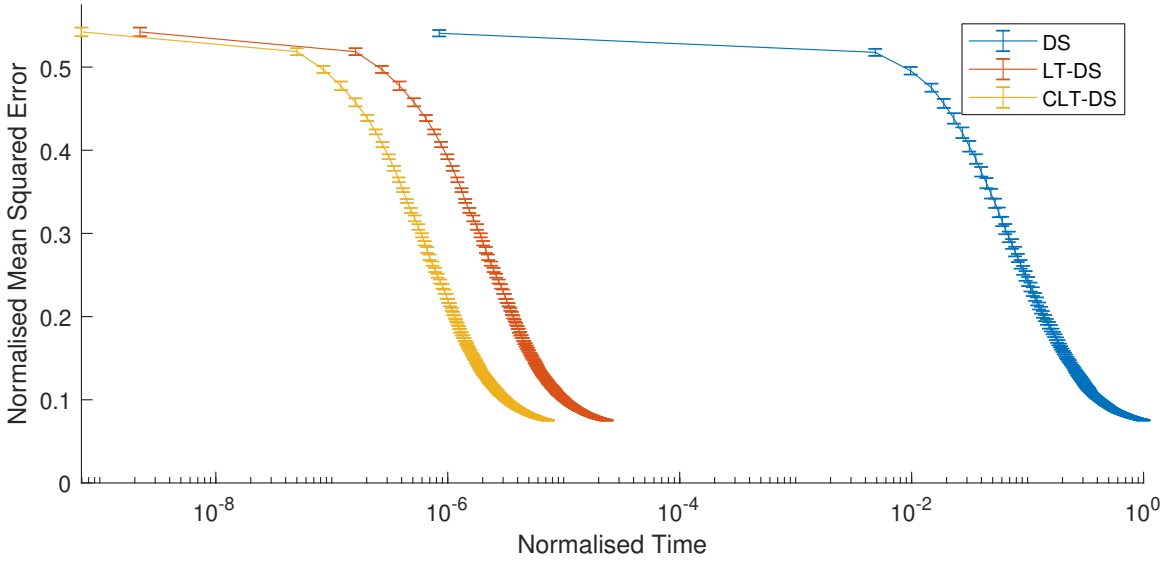


Fig. 5.37 Convergence of direct search and linear time direct search for a simulated 1024×1024 pixel 2^8 phase level spatial light modulator.

$$\Delta E_{MSE}(G,H) = \left|G_{x,y} - H_{x,y} - \Delta H_{x,y}\right|^2 - \left|G_{x,y} - H_{x,y}\right|^2 \tag{5.50}$$

which runs in constant $O(1)$ time whereas on the replay side the update runs in $O(N_x N_y)$ time. This error calculation can be incorporated into the direct search algorithm - Figure 2.14e - to give linear time direct search (LT-DS). Running the LT-DS algorithm gives the performance graph shown in Figure 5.37. Target amplitudes are given by the *Mandrill* test image and target phases are given by the *Peppers* test image as shown in Figure 5.19. With 1024×1024 pixel test images this gave a $\approx 50,000\times$ speed up for the DS algorithm. In this implementation, this corresponds to a runtime of 12 milliseconds as opposed to approximately 10 minutes. Similar results are seen for simulated annealing. Due to the amplitude and phase constraint on the target, however, convergent reconstruction quality is extremely poor. This is traditionally solved by using an RoI, a topic returned to in Section 5.6.7.

It is important to note that, provided the random number generators have the same seed, the hologram given by LT-DS is identical in every way to the hologram provided by DS. The only difference is the Fourier plane on which calculation occurs and the resulting orders of magnitude speed-up. Also worthy of note is the normalisation of the values of the hologram here to give a mean of unit energy per pixel on SLM and replay field sides, with a resulting normalisation effect on the MSE.

Result: The change in mean squared error of a phase sensitive hologram due to a single pixel change can be found in constant $O(1)$ time.

5.6.4 Effect of Independence

Section 5.6.3 used (5.50) to reduce the computation required for DS but maintained the use of the search approach. There are cases, such as when an RoI is taken into account (Section 5.6.7), where a search approach is necessary, but for the RoI-free case discussed here, search syntax is not required at all. Instead, noticing that the effect on the MSE of changing a single pixel is independent of the other pixels. This means that the search element can be removed altogether, instead independently assigning values to each individual pixel. This is important as it allows parallelisation of the algorithm for multi-core devices. The performance improvement obtained in this way over the sequential version is also shown in Figure 5.37 and has been termed concurrent LT-DS or CLT-DS. The workstation used had a Intel® i7-9900K CPU, overclocked to 5.0GHz with 64GB of 4000MHz DDR4 ram and a RTX 2080TI GPU.

Result: The change in mean squared error of a far-field phase sensitive hologram due to a single pixel change is independent of the effect of other pixels.

5.6.5 Realistic SLM Constraints

The form of (5.49) is a linear minimisation problem and is solvable analytically for a range of modulation regimes. This dependency on the properties of the modulator requires investigation of the case of phase and amplitude modulating devices separately.

Phase Modulating

If a phase modulating device is assumed where $H_{x,y}$ is confined to the complex circle with magnitude given by the incident illumination $I_{x,y}$ then (5.50) becomes

$$\begin{aligned} \text{minimise} \quad & \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \left| G_{x,y} - H_{x,y} \right|^2 \\ \rightarrow \quad & \Phi_H = \Phi_G \end{aligned} \tag{5.51}$$

where Φ_G and Φ_H correspond to the phase vectors of G and H.

Result: When aberration and replay field RoIs are neglected, the lowest possible mean square error is achieved for a far-field phase hologram when the phase is equal to the inverse transform of the target replay.

Amplitude Modulating

If an amplitude modulating device where $H_{x,y}$ is assumed to be confined to $\left|H_{x,y}\right| \geqslant 0$ and $\Phi_H = 0$ then (5.50) becomes

$$\text{minimise} \quad \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \left|G_{x,y} - H_{x,y}\right|^2 \quad \rightarrow \quad H = \Re(G) \tag{5.52}$$

Result: When aberration and replay field RoIs are neglected, the lowest possible mean square error is achieved for a far-field amplitude hologram when the SLM amplitude is equal to the real part of the inverse transform of the target replay.

5.6.6 Fresnel Holograms, Aberration Correction and 3D

The Fresnel transform used for generating Fresnel holograms is equivalent to the Fourier transform with the addition of a *quadratic phase factor* as in

$$R_{u,v} = \underset{\text{Fresnel}}{\mathscr{F}}\{H_{x,y}\} = \underset{\text{Fourier}}{\mathscr{F}}\{H_{x,y}\Phi_{\text{Fresnel}}\} \tag{5.53}$$

where $\Phi_{\text{Fresnel}} = \exp\frac{i\pi}{\lambda z}(x^2+y^2)$. It can be seen that the Parseval theorem remains applicable here; (2.68) and (5.49) remain equivalent and the results of Section 5.6.5 remain valid with the addition of an additional phase term.

In fact, for any input phase term dependent only on x and y the equivalence of (2.68) and (5.49) can be asserted. This includes the family of Seidel aberrations.

While this section discusses the linear-time algorithm in the context of 2D holograms, it is equally applicable to 3D holograms generated by means of *Fresnel slices* or the layer based technique.

5.6.7 Incorporating a Region of Interest

The reconstruction quality obtained for complex-valued target fields using the techniques above is often extremely poor, but this is not due to the choice of algorithm. Instead, this is because the problem is over-constrained. One solution that is widely used is to only require a portion of the replay field to match the target image, with the remainder of the replay field being free to take on any value. Mathematically this gives the definition of RoI mask $M_{u,v}$ where $M_{u,v} = 1$ in the RoI and $M_{u,v} = 0$ otherwise. (2.68) can then be rewritten as

$$E_{MSE}(T,R) = \frac{1}{N_x N_y} \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} M_{u,v} |T_{u,v} - R_{u,v}|^2 \tag{5.54}$$

Unfortunately. Parseval's theorem - (2.37) - no longer allows this to be moved to the SLM side, as Parseval's theorem only holds true if all of space is considered, instead of only a subregion of space.

Instead, an alternative technique for incorporating an RoI into a linear time algorithm allows rewriting (5.54) to give

$$\begin{aligned}
E_{MSE} &= \frac{1}{N_x N_y} \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} |M_{u,v}T_{u,v} - M_{u,v}R_{u,v}|^2 \\
&= \frac{1}{N_x N_y} \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} M_{u,v}T_{u,v}\overline{M_{u,v}T_{u,v}} - M_{u,v}T_{u,v}\overline{M_{u,v}R_{u,v}} - \\
&\qquad \overline{M_{u,v}T_{u,v}}M_{u,v}R_{u,v} + M_{u,v}R_{u,v}\overline{M_{u,v}R_{u,v}} \\
&= \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} F_{x,y}\overline{F_{x,y}} - F_{x,y}\overline{(L*H)_{x,y}} - \\
&\qquad \overline{F_{x,y}}(L*H)_{x,y} + (L*H)_{x,y}\overline{(L*H)_{x,y}} \\
&= \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} F_{x,y}\overline{F_{x,y}} - F_{x,y}\overline{K_{x,y}} - \overline{F_{x,y}}K_{x,y} + K_{x,y}\overline{K_{x,y}}
\end{aligned} \tag{5.55}$$

where

$$L \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\rightleftarrows}} M, \quad F \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\rightleftarrows}} M \cdot T, \quad K \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\rightleftarrows}} M \cdot R$$

$F_{x,y}$ behaves similarly to previously and single pixel updates can be determined in $O(1)$. $K_{x,y}$ corresponds to a convolution though, and cannot be evaluated as easily. Fortunately, while convolution is an $O(N_x^2 N_y^2)$ problem, changing a single pixel of a convolution can be somewhat optimised. The convolution term of (5.55) is given for any pixel x', y' by

$$K_{x',y'} = \sum_{a=0}^{N_x-1} \sum_{b=0}^{N_y-1} L_{a,b} H_{x'-a,y'-b} \tag{5.56}$$

Recognising that L is only non-zero for a handful of pixels, this can be calculated in $O(n)$ where n is the number of pixels where $L \neq 0$.

$$K_{x',y'} = \sum_{L \neq 0} L_{a,b} H_{x'-a,y'-b} \tag{5.57}$$

A change in a single pixel x, y of value $\Delta H_{x,y}$ then causes a difference to the convolution at pixel x', y' of

$$\Delta K_{x',y'} = L_{x'-x,y'-y} \Delta H_{x,y} \tag{5.58}$$

Incorporating this back into the MSE equation, the following update step can then be defined.

$$\Delta E_{MSE} = \frac{1}{N_x N_y} \sum_{x'=0}^{N_x-1} F_{x,y} \overline{\Delta K_{x,y}} - \overline{F_{x,y}} \Delta K_{x,y} + \Delta K_{x,y} \overline{K_{x,y}} + K_{x,y} \overline{\Delta K_{x,y}} + \Delta K_{x,y} \overline{\Delta K_{x,y}} \tag{5.59}$$

This can be incorporated into the DS algorithm of Figure 2.14e. Any given mask $M_{u,v}$ can be given to an arbitrary degree of accuracy by $\mathscr{F}\{L\}$ though in practice if L is non-zero for more than a few points a change of mask or an alternative approach is recommended.

To demonstrate this in action we take the case of L being non-zero only at a selected 45 points out of a 512×512 image. This leads to a mask function similar to Figure 5.38 with associated figures.

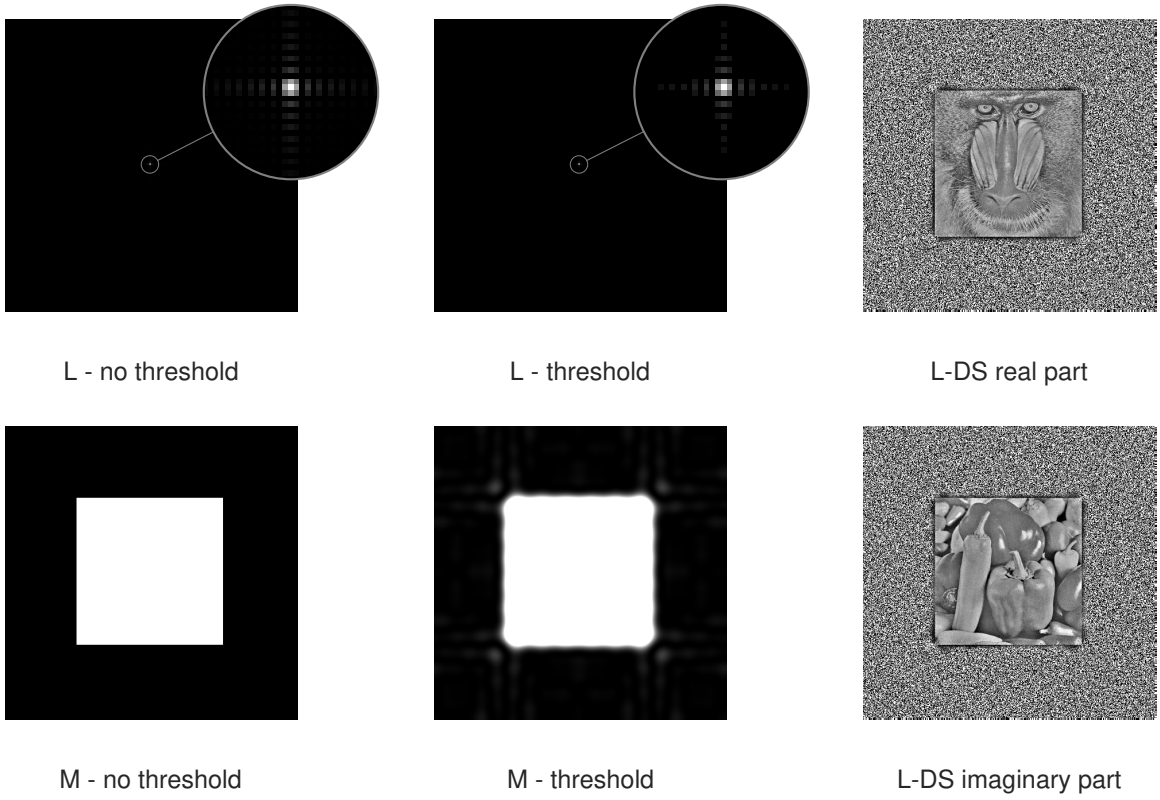


Fig. 5.38 Mask and inverse transform of mask without thresholding (left) and with thresholding (centre) Reconstruction real and imaginary parts for LT-DS are shown right.

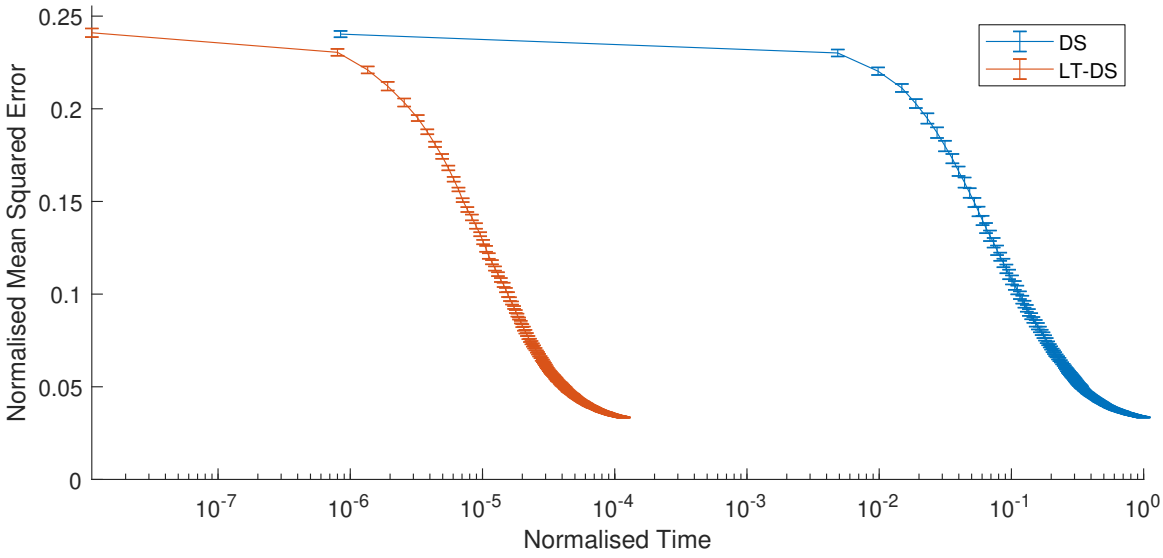


Fig. 5.39 Performance of direct search and linear time direct search for a simulated 1024×1024 pixel 2^8 phase level spatial light modulator with mask region thresholded at 45 points.

The quality of the mask in Figure 5.38 depends on the threshold value chosen. For many simple masks, over 90% of the power in the mask can be captured by only a few points in L. This corresponds to a slight re-weighting of MSE priorities due to differences in value of M.

The performance scales linearly with the number of points in L. For the images in Figure 5.38 with L thresholded to 45 points, we see the performance shown in Figure 5.39 with identical normalisation to that in Figure 5.37. The speed improvement when compared to Figure 5.37 is lower, however, due to higher number of calculations per iteration but is still $10,000\times$ faster than the traditional DS approach. In this implementation this is equivalent to a 900 millisecond runtime as opposed to approximately 10 minutes.

As in Section 5.6.3, the hologram generated using this approach is identical to generating a hologram using DS with mask function M provided the same random number generator seeds are used in both cases.

5.6.8 Further Work

The work described so far is applicable in the case where both the phase and amplitude of the replay field are to be controlled. The progress made prompts the obvious question of whether this linear time technique can be applied to phase insensitive holograms where the error is given by

$$E_{MSE,pi} = \frac{1}{N_x N_y} \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} [|T_{u,v}| - |R_{u,v}|]^2 \tag{5.60}$$

Clearly this problem is non-linear so a best possible solution is improbable. The author believes, however, that the techniques of this paper should allow a similar movement of an error metric to the SLM side, but have so far been unable to implement this.

5.6.9 Summary

This section has presented a new approach to generating holograms for two-dimensional phase sensitive replay fields. The discussed algorithm relies a judicious use of the Parseval's theorem, allowing the phase-sensitive MSE error metric to be calculated from the field in the

SLM plane. This allows search algorithms such as SA and DS to run without the need for repeated Fourier transforms, providing a significant acceleration in execution time. Whereas one iteration of a more traditional DS algorithm has a computational cost of $O(N_x N_y)$, iterations of the new proposed implementation have a computation cost as low as $O(1)$. This performance boost is particularly marked for high-definition holograms. For example, with the Tokyo 2020 Olympics being shown in 8k (7680×4320) resolution, the expected performance improvement is over 1 million times faster. The algorithm has been presented for Fraunhofer holograms, but has been shown to be equally valid for Fresnel holograms. Conclusions have been drawn for common modulation schemes. An equivalent approach for a phase-insensitive MSE error metric has not been found, but it is felt that further work can address this.

5.7 Relevance

This chapter has introduced 5 new algorithms: Sorted Pixel Selection, Single-Transform Time-Multiplexed, Holographic Predictive Search, Sympathetic Quantisation and Linear Time. These algorithms are all valuable in their own right but offer varying applicability to the particular aim of this project, additive manufacture. This section begins be introducing selection criteria for algorithm selection in 2D CGH and then continues to discuss the particular place that these 5 algorithms fill in this ecosystem. In addition, Appendices A-D offer further discussion of algorithmic elements.

5.7.1 Selection Criteria

The choice of algorithms for CGH is involved requiring detailed knowledge of the application and SLM used. The primary considerations include the modulation capabilities of the SLM, the form of the target images and whether the target is phase sensitive or phase insensitive. While the best results can be achieved only with expert insight. a simplistic choice method is shown in Figure 5.40. This covers some of the more common algorithms introduced in Section 2.4 but fails to account for the many intricacies of holographic optical systems.

From Figure 5.40, three broad categories of algorithms can be determined.

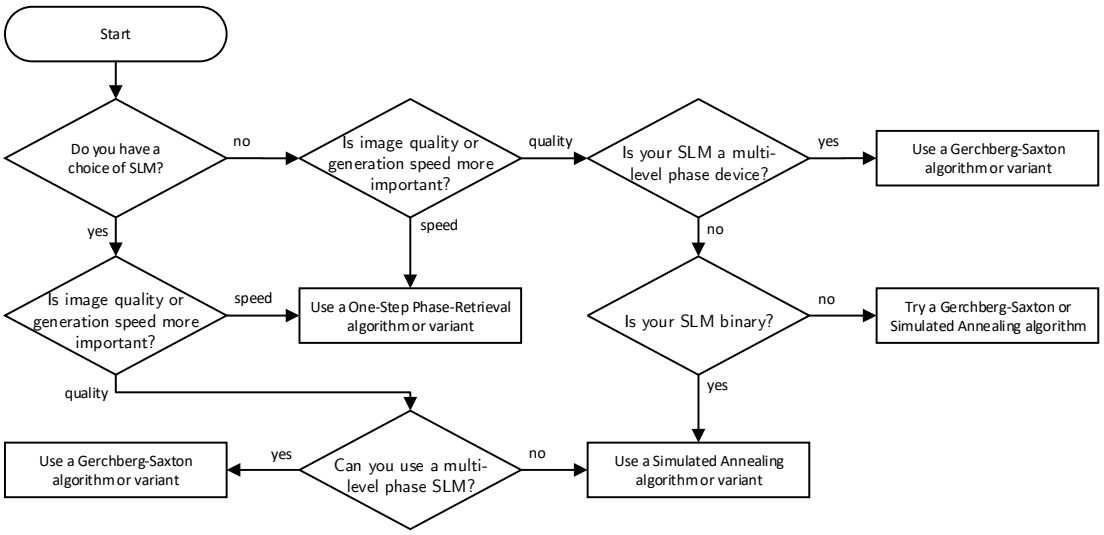


Fig. 5.40 Basic decision process for algorithm choice when designing a holographic system.

Firstly, iterative algorithms such as Gerchberg Saxton, Section 2.4.3, which operate by repeatedly transforming a whole image in order for it to converge to an optimum solution. These are typically operated on multi-phase devices as they offer the 'smooth' variation in values conducive for solution finding. Notable these largely don't work, without significant problem transformation, in the case of binary devices or phase sensitive replay fields.

Secondly, search algorithms such as direct search or simulated annealing, Section 2.4.4, which operate using a trial and error approach switching pixels or groups of pixels and accepting the result on the basis of some acceptance function. While these work for the problems covered by iterative algorithms, they are orders of magnitude slower. As a result, these algorithms are largely targeted at binary devices and/or problems with a phase sensitive replay field.

Thirdly, time-multiplexed algorithms such as one-step phase-retrieval, Section 2.4.5, which operate by time averaging a large number of independent sub-frames. These are focussed exclusively at real-time display applications with a phase insensitive replay field and typically on binary devices.

- **Sorted Pixel Selection** is an method of optimising search algorithms including DS and SA.

- **Single-Transform Time-Multiplexed** is a time-multiplexed algorithm that offers significant improvements in run-time and quality over its competitors.
- **Holographic Predictive Search** is a predictive model for search algorithms that allows them to operate significantly faster at the expense of increased complexity.
- **Sympathetic Quantisation** is a hybrid algorithm, operating on multi-phase devices but in the context of single frame generation.
- **Linear Time** is a search algorithm.

5.7.2 Relevance to Additive Manufacture

Depending on the designed system, all three classes have applications to additive manufacture. During the early stage of work on the PhD it was intended to operate with an array of micro projectors based on binary SLMs. This gave rise to the STTM algorithm, Section 5.3, a significant improvement on the existing time-multiplexed algorithms. In particular, STTM offered guarantees on the convergence of total error unlike its competitor, OSPR. While this was tested on binary devices, the micro projector project is still ongoing an the time of submission.

As a result, the resin demonstrator, Section 6.2, used a multi-phase device. For pre-generated frames, this originally used SPS, Section 5.2, combined with an SA, Section 11, algorithm. This was later replaced by HPS, Section 5.4, due to its improved performance and reliability.

While these frames remained the gold-standard in quality, for real-time and near-real-time applications, Gerchberg-Saxton was used before being replaced by SQ, Section 5.5. SQ offered order magnitude performance increases at the expense of worse noise for any given frame. Importantly, SQ doesn't phase randomise. For display applications this causes undesirable edge-enhancement but for AM exposures lasting several seconds, this is not an issue and can be shown to average out. Phase randomisation is known to cause significant sub-pixel noise and the use of SQ significantly reduces that while improving the feature accuracy and sharpness of edges[7]. Unless otherwise stated, every figure shown in Chapter 6 uses holograms generated using the SQ algorithm.

The only algorithm here that was not directly used during development of the AM demonstrator in this project was LT. This was originally conceived for fibres with the end goal that complex

[7]This is the subject of a forth-coming paper on interstitial noise in holograms

control of the phase input to a fibre would allow control of the output. It was originally intended that this could be then incorporated into a more traditional extrusion or jetting AM progress that would allow for customised profile delivery without the use of galvoscanners.

5.8 Conclusion

This chapter has introduced the four new algorithms used for hologram generation for AM during the project in Sections 5.2 to 5.5 with Section 5.6 introducing an algorithm of potential future interest. Section 5.7 then discussed how these algorithms fit into the infrastructure of holographic algorithms and their relevance to the wider AM project.

Chapter 6

Demonstrators

Chapter 6

Demonstrators

6.1 Introduction

While the project to build a holographic metal powder system is ongoing, a low power practical demonstration can be achieved with a holographic projection system included in a resin stereolithography (SLA) printer process. This uses many of the areal techniques discussed without requiring expensive, high-power optical components. The design, implementation and results of this are discussed in Section 6.2. Section 6.3 continues to discuss the progress, to date, of work on a high power metal powder demonstrator.

6.2 Resin Demonstrator

6.2.1 Introduction

In order to demonstrate the power of holographic stereolithography (HSLA), the Prusa SL1 was chosen as a comparison devise. The Prusa SL1 is a resin printer with specifications summarised in Table 6.2. With the exception of build volume, the resin printer was required to at least match the performance of the Prusa machine.

6.2.2 Mechanical Design

Mechanically the optical system is designed for exposure upwards into a resin tank with a replacable transparent FEP film base. A mechanical stage with custom platform is moved vertically through the resin with negative pressure drawing the viscous resin beneath the platform as it is raised. The system was designed and tested for layer heights of between 10µm and 100µm.

For the resolutions required, basic steady state thermal models and vibration analysis carried out in Fusion 360 determined that the combined thermal and vibration error expected in the lab environment was equivalent to less than 3µm. This would have been reduced to under 1µm had a stiffer laser mount been used. As this is significantly smaller than the printers target 50µm resolution, thermal and vibration error have been ignored.

The basic design of the printer is annotated in Figure 6.2 with end result shown in Figure 6.1.

6.2.3 Optical Design

The optical system designed uses a SureLock™ LM Series 405nm, 40mW laser diode - Table 6.3 - and a Jasper JD8714 SLM - Table 6.1. This is configured in with a incident beam 15° to the SLM perpendicular and the associated aberrations compensated for in the code. An overview schematic is given in Figure 6.3.

Light is emitted from the 405nm laser before being focussed through a point by a 10mm aspheric lens. This focal point becomes the projector's limited point source and a 15mm pinhole is used to spatially filter the beam at this point. A $\lambda/2$ wave plate is used to rotate the major axis of the 100:1 polarisation ellipse to lie along the SLM's optical axis[1]. A 150mm plano-convex lens is used after beam expansion before the light is reflected off of the SLM.

The beam is then passed through a polarizing beam splitter (PBS) where the s-polarised light is transmitted and captured by a beam dump and the p-polarised light is reflected. The light is then passed through a longpass 425nm dichroic mirror before being expanded using a 20× objective and projected onto the surface of the resin. The dichroic mirror serves to allow for an integrated microscope imaging system that can be used to measure the process in real time.

[1]The Jasper JD8714 has an optical axis oriented at 45° degrees to the SLM major axis.

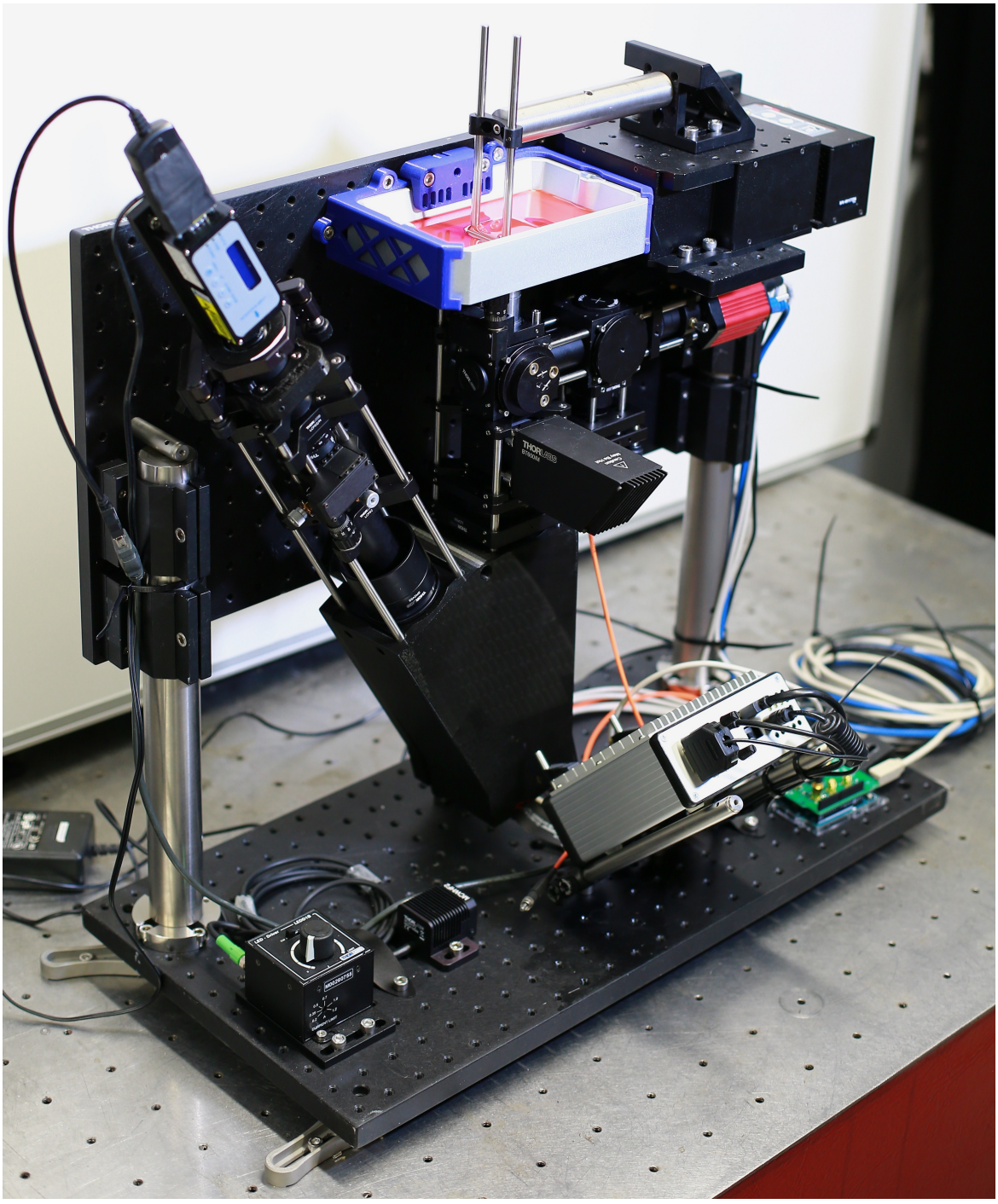


Fig. 6.1 Resin printer photograph

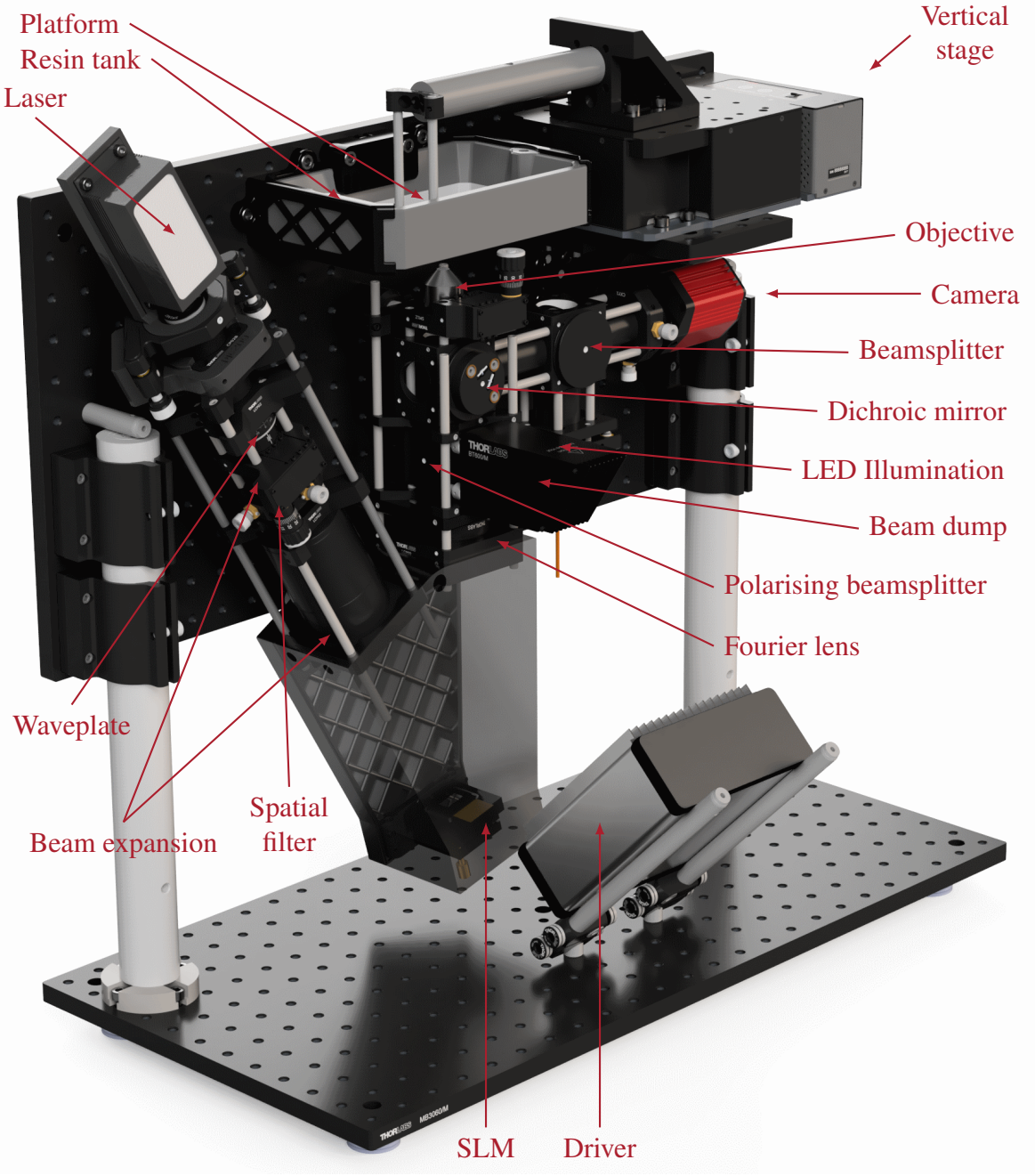


Fig. 6.2 Resin printer configuration

Table 6.1 Jasper JD8714 specification

Parameter	Specification
Package Size	31.00mm by 24.00mm
Active Area	15.6mm by 9.2mm
Pixel Pitch	3.74µm
Fill Factor	89%
Resolution (active)	4096 pixels by 2400 pixels
Resolution (addressable)	4160 pixels by 2464 pixels
Modulation Type	8-bit Phase

Table 6.2 Prusa SL1 specification

Parameter	Specification
Illumination	990 mW/m^2
Wavelength	405nm
Print area size	$120 \times 68 \times 150$ mm
Resolution	47 µm
LCD	5.5”, 2560×1440 pixels
Layer height	0.025mm-0.1mm
Speed	6 seconds per layer, independent of layer size

Table 6.3 SureLock™ LM-405-40 specification

Parameter	Specification
Output Power	40mW
Center Wavelength (vacuum)	405nm$\pm$0.5nm
Linewidth	160MHz
Bandwidth	87fm
Beam Size (FWHM)	0.8mm by 0.4mm
Polarization Ratio	100:1
Polarization Orientation	TE

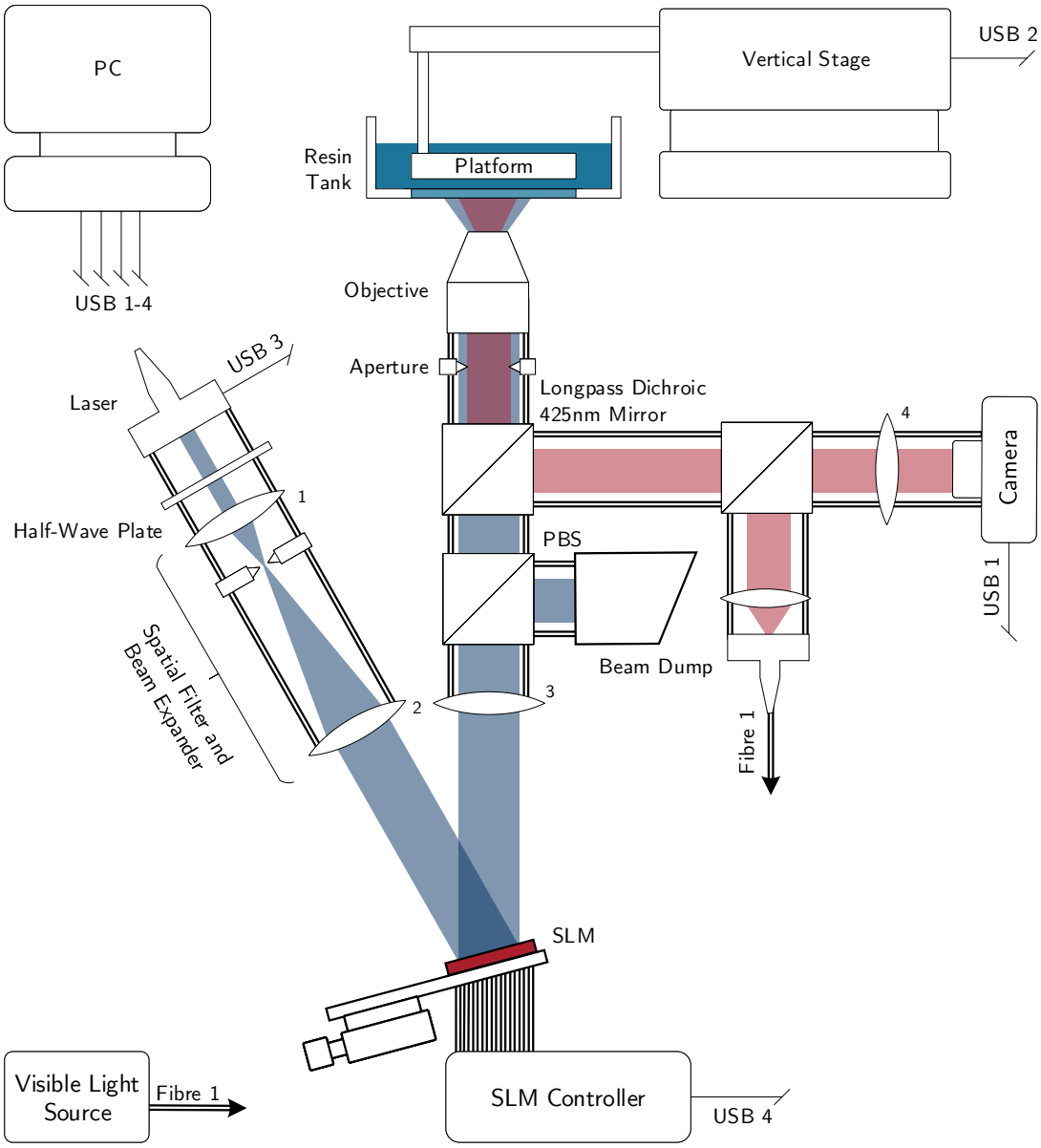


Fig. 6.3 Resin demonstrator schematic

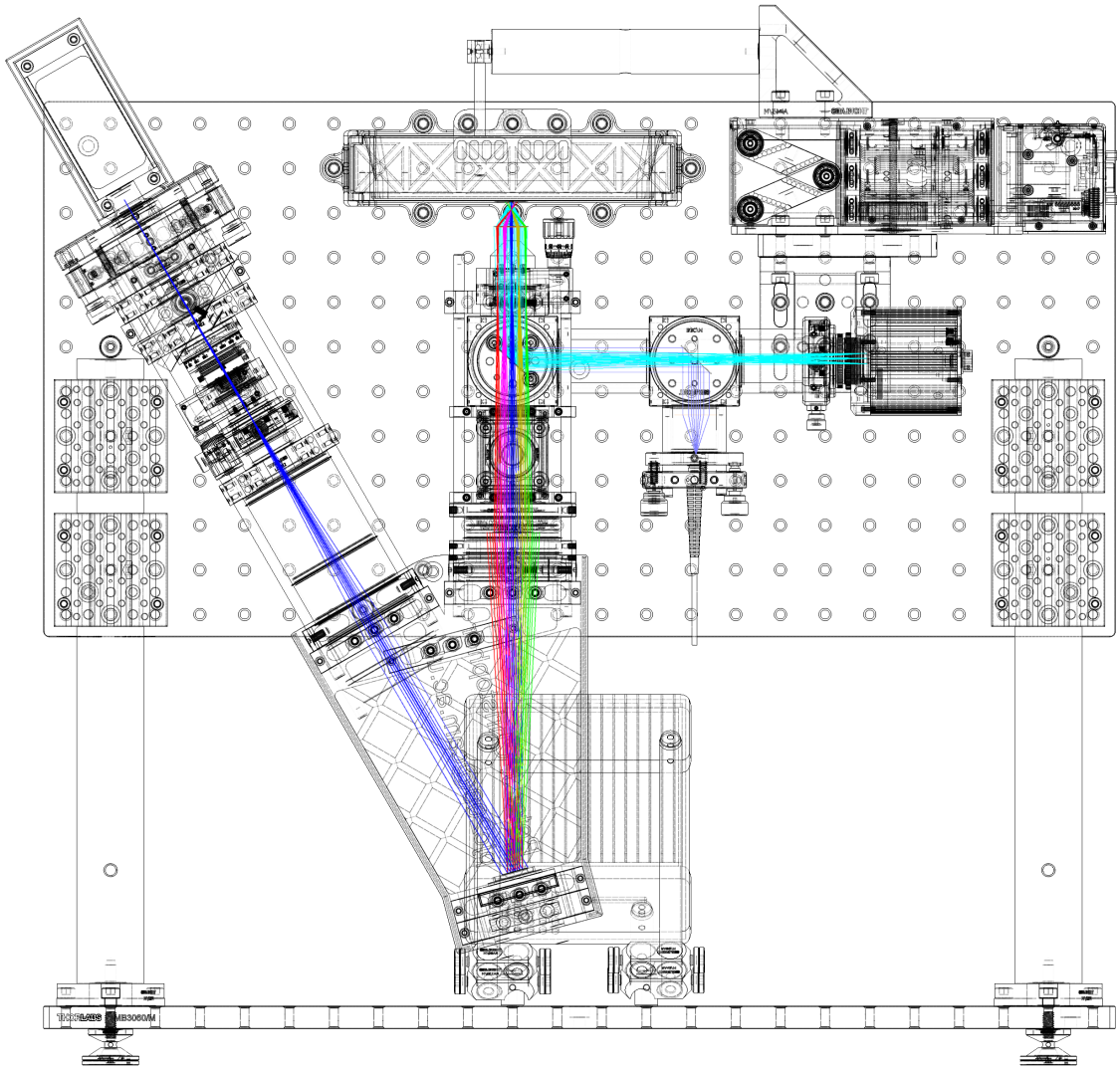


Fig. 6.4 Optical design for resin printer

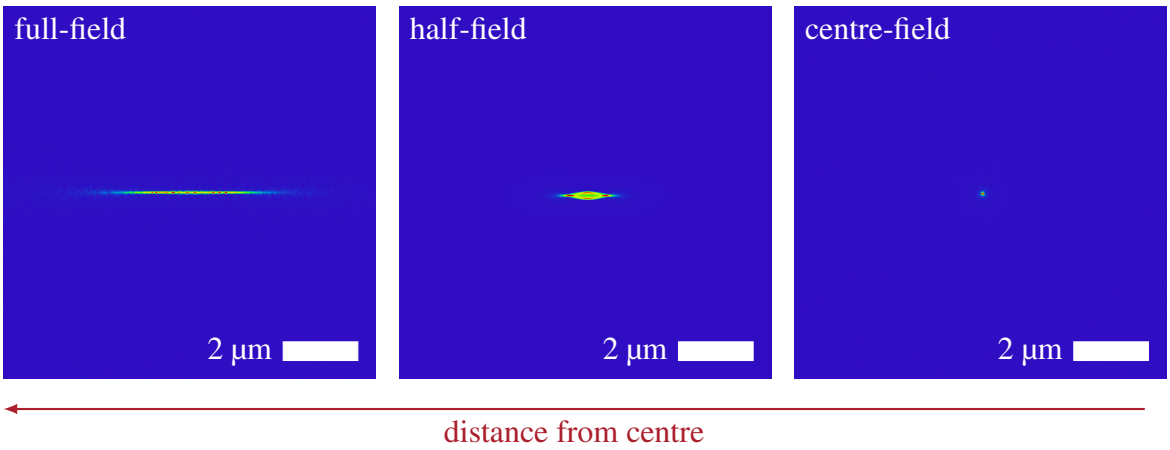


Fig. 6.5 Spot diagrams across the replay field as calculated in Zemax OpticStudio

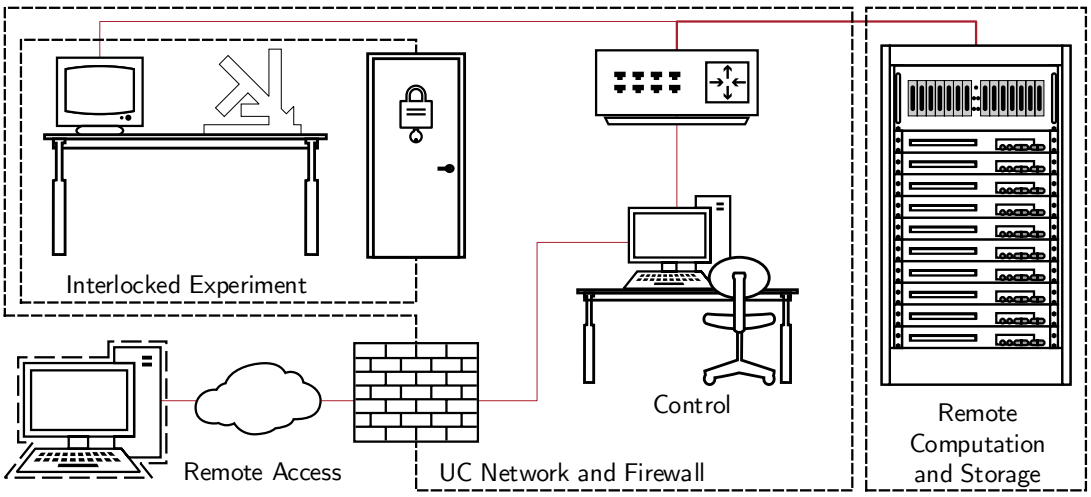


Fig. 6.6 Distributed experimental control

A Zemax OpticStudio Non-Sequential model, Figure 6.4, was used to test the configuration with a beam path. The full prescription and model files are linked in Appendix I. From Section 2.2.7, (2.54), the replay field has size 99.4×58.2mm before the objective with a diffraction limited spot diameter of 24µm. This is significantly larger than the ray traced spot sizes shown in Figure 6.5 and shows that the system is diffraction limited.

Calibration and Aberration Correction

The exposure pattern is tested by exposure on a 3.54µm pixel pitch camera both with and without the microscope objective. By showing a sequence of target images and recording the result, a deconvolutional alignment system is used to determine a sequence of correction functions that are applied to the SLM to correct for distortions and aberrations. By using discrete cosine transforms (DCTs) instead of the more traditional Zernike basis, it is possible to retain some of the symmetry properties that are utilised by the hologram generation. A linear canonical transform (LCT) method is used to allow for a linear solution to what is usually a non-linear problem. The full approach is being prepared for submission to a journal.

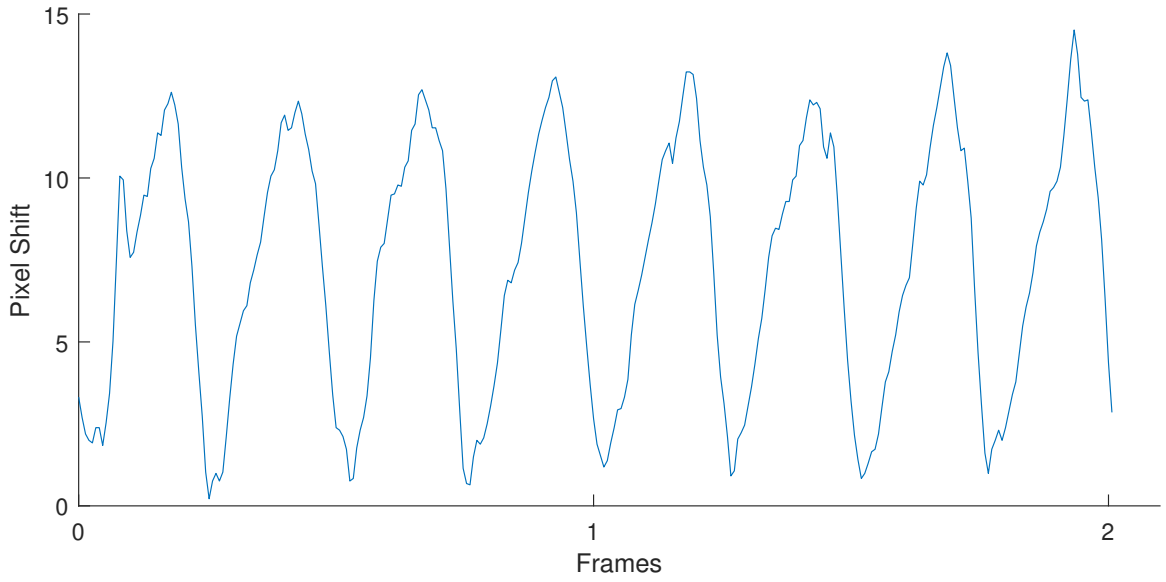


Fig. 6.7 SLM phase drift

Computer Generated Holograms

The holograms for the printer are generated using the Soft Sympathetic Quantisation (SSQ) method introduced in Section 5.5. This allows for faster runtimes as well as improved noise performance. This also allows for easy addition of the aberration correction mentioned above and the strictly even/odd symmetry of the DCT allows for the corrections to be applied to the hologram after generation rather than during. This allows for caching and greater separation of hologram generation and experimental control. This approach is also being prepared for submission to a journal.

Software Design

The software for the resin printer was coded in Matlab, C# and Cuda C++. Based on HoloGen, Chapter 4, this is responsible for generating holograms, automated alignment, photogrammetry and component interfacing.

The slicing code is currently based on Autodesk's Netfabb application with a custom parameter set. It is intended that there be a custom slicing application build for the powder demonstrator.

A brief overview of the structure is available in Appendix H. Unlike HoloGen, the code is not yet publicly available as it currently depends on unpublished algorithms.

Importantly, the software is designed for distributed generation, Figure 6.6, where the computationally expensive hologram generation is executed on a server machine with the lab machine responsible for component interfaces. This can be controlled remotely using a Windows Web Services (WWS) API service model interface.

The package also includes an API link to the Zemax OpticStudio model used for design in Section 6.2.3 to directly inform the initial alignment which is then used as a seed for the calibration and aberration correction of Section 6.2.3.

6.2.4 SLM Behaviour

The Jasper SLM used exhibited significant phase instability and drift issues. For a camera synchronised to the SLM VSync pulse, stability is excellent and showed drift of less than $1/10^{\text{th}}$ of a replay field pixel. Pixel here, being taken as an arbitrary unit of measurement equal to the replay field size divided by the resolution. When the camera was not synchronised, however, drift was high with the replay field drifting the equivalent of 10-15 replay field pixels over the course of a single frame. The total average frame drift relative to the start of the frame is shown in Figure 6.7[2].

The reason for this is the pulse width modulation (PWM) occuring within the SLM. While it is possible to rewrite the PWM file, this comes at the expense of lowering the number of achievable phase levels, tests suggested that, to ensure a stability with equivalent drift of less than a pixel in the replay field, only 8-10 phase levels could be used. It is suggested that this is due to the JD8714 being Jasper's first 4K device, and that clock speeds suitable for earlier HD devices were insufficient for a device with roughly four times the pixels. Given that Jasper never released the JD8714 prototype, it is reasonable to assume that this issue was the cause. Particularly as Jasper moved away from the FPGA based driver in the JD8714 to a higher frequency ASIC chip in later models.

The drift is compensated for in a number of ways. Firstly, by using a frequency balancing regime in the phase term of the hologram. Provided the phase terms are sufficiently independent of each other and approximately uniformly distributed, the effect on the

[2]See Appendix I for a composite video showing the drift over the course of a single frame.

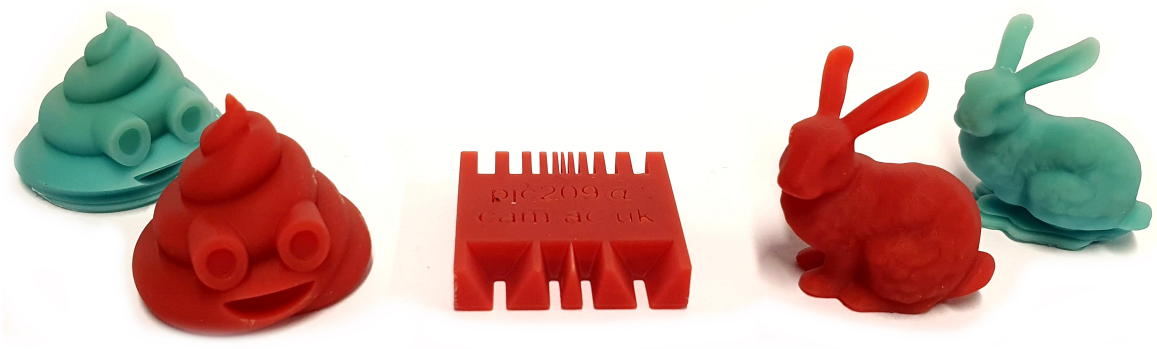


Fig. 6.8 Resin printer results showing holographic additive manufacture (red) with Prusa SL1 comparison (blue). Prusa parts retain the print platform

hologram is largely confined to a scaling term with the remainder of the power shifted into the zero order.

Secondly by using fewer phase levels with a custom bit plane map. This increases the spatial noise which in turn is compensated for by the temporal averaging of the independent frames.

Thirdly by internally offsetting the boundaries of the exposure regions and proportionately overexposing the resin. This is acceptable as that limited overexposure of the resin does not change the cure behaviour. The downside of this is an increased minimum feature size. This remains below the limit of feature size the resin can support, however, so is not a problem.

Aside from the drift, the SLM PWM cycle also caused significant variation in phase sensitivity over the frame cycle, causing a zero order that was impossible to remove. While off-axis configurations were considered this would have caused a significant drop in far-field resolution and required a more powerful laser to meet exposure requirements. Instead, the zero order was reduced as far as possible and allowed to remain. As the printer must always expose a zero order and the samples shown here are designed to accommodate this. Were this ever to be used in production, a better performance SLM would be used without this issue.

6.2.5 Results

The demonstrator produced excellent results as shown in Figure 6.8. Here the HAM demonstrator printed the parts at the centre in red and the Prusa SL1 machine produced the outer parts in blue. Note that the Prusa parts still contain a platform which accounts for the slight differences at the base of the parts.

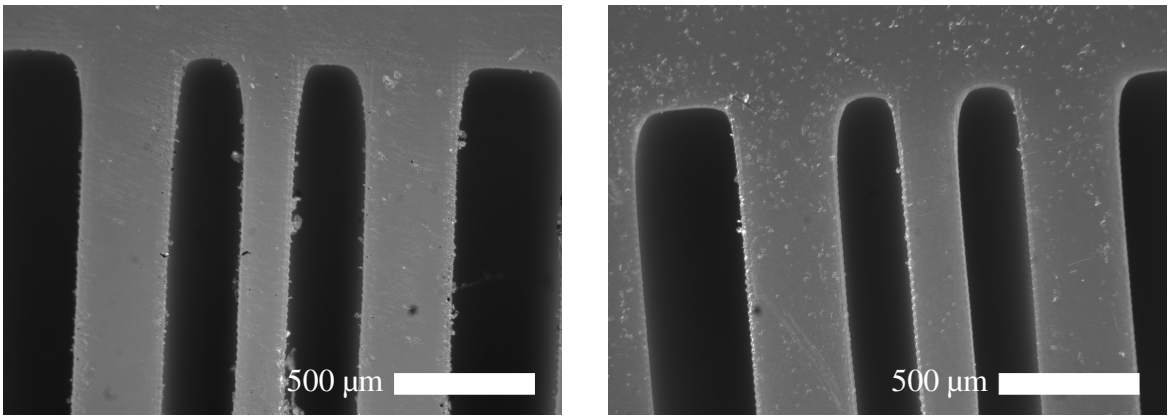


Fig. 6.9 Close-up microscope images of the resin prints showing the Prusa SL1 print (left) and resin demonstrator (right).

The real-time imaging allows for a measurement of physical accuracy which can be later confirmed using an optical microscope. Figure 6.9 shows an optical microscope image of a test part with 150µm thinnest wall and 250µm thinnest cavity as printed on the Prusa SL1 (left) and HAM demonstrator (right). It can be seen that both systems maintain the 50µm horizontal resolution specified for the SL1.

The HAM system did experience some issues with lift off as the system lacked the tilting bed used by the Prusa machine but this was solved by using an over compensation of the stage between each layer. When the stage is to be lifted by a layer height, it actually travels upwards by three layers before being lowered down to the correct height. This has the issue of reducing the negative pressure pull on the part during exposure and curing and relies on the effects of buoyancy and high resin viscosity to keep the part pushed into the bed.

6.2.6 Further Work

While the HAM system is intended primarily as a demonstrator, there is scope for further investigation. This section has not explored the lower limits of achievable resolution, nor the opportunities for dynamic and super resolution imaging. It is intended to explore these further at a later date.

The achievable resolution, in important as, unlike LCD displays, the size of the produced hologram is limited only by the diffraction limits of the light rather than the physical pixel size. This allows for the resolution of the projected area to be scaled purely by changing the objective lens. This should allow for printing at smaller scales than seen previously.

Additionally, the hologram can be used for 'dynamic resolution' capabilities. The positioning of the replay field sampling points are independent of the positioning of the SLM pixels. This allows for oversampling and undersampling in different regions of the image depending on feature density. The precise limits of how far this effect can be taken remain to be explored, however.

While it is not anticipated that holography will replace current projection systems for current resin printers, it is anticipated that the increased resolutions achievable will make it a desirable approach for printing at below the micron scale in applications such as microfluidics.

6.2.7 Summary

This section has introduced the holographic additive manufacture (HAM) demonstrator for printing in resins. The prototype system was compared to a mid-range commercial stereolithography process with comparable performance on all metrics. The next section continues to discuss the work done so far on developing the high-power powder prototype.

6.3 Powder Demonstrator

6.3.1 Introduction

It was not possible to complete the plastic/metal powder demonstrator during the period of the PhD. As follow-on funding has been acquired to complete the project, this section discusses the progress so far as well as the work packages to be carried out in the remaining two years of the follow-on Engineering and Physical Sciences Research Council (EPSRC) grant.

6.3.2 Concept

The planned powder system has a schematic as shown in Figure 6.10. A high-load vertical stage moves the build platform vertically. The platform has embedded heating elements for pre-heating of the powder. Unlike standard selective laser sintering/melting (SLS/SLM)

Table 6.4 Santec SLM-300 specification

Parameter	Specification
Package Size	122.0mm by 122.0mm by 38.3mm
Power Handling	200W/cm^2 @ CW 1064nm
Active Area	15.36mm by 9.60mm
Pixel Size	7.80µm
Pixel Pitch	8.0µm
Fill Factor	95%
Resolution	1920 pixels by 1200 pixels
Modulation Type	10-bit Phase
Frame Rate	60-120Hz

Table 6.5 SPI SP-200C-0001 specification

Parameter	Specification
Output Power	200 W
Mode	CW
Polarisation	Random
Wavelength	1090nm $\pm$ 5 nm
Bandwidth	4nm

machines, the powder is carried on a trolley on top of the spreader rather than dispensed by a parallel piston, Figure 1.1b. This allows for a reduction in enclosure volume and greater flexibility in powder arrangement at the expense of slower operation.

The platform is illuminated by a simple holographic projector consisting of a laser source and collimating optics, SLM and Fourier lens. The initial demonstrator will incorporate only one projector with planned production systems using multiple projectors covering the same area. A centrally aligned vision system will image the bed in the visible and at thermographic wavelengths of 8-12µm.

The system is enclosed in an argon filled enclosure to prevent oxidisation and the exhaust is filtered before extraction.

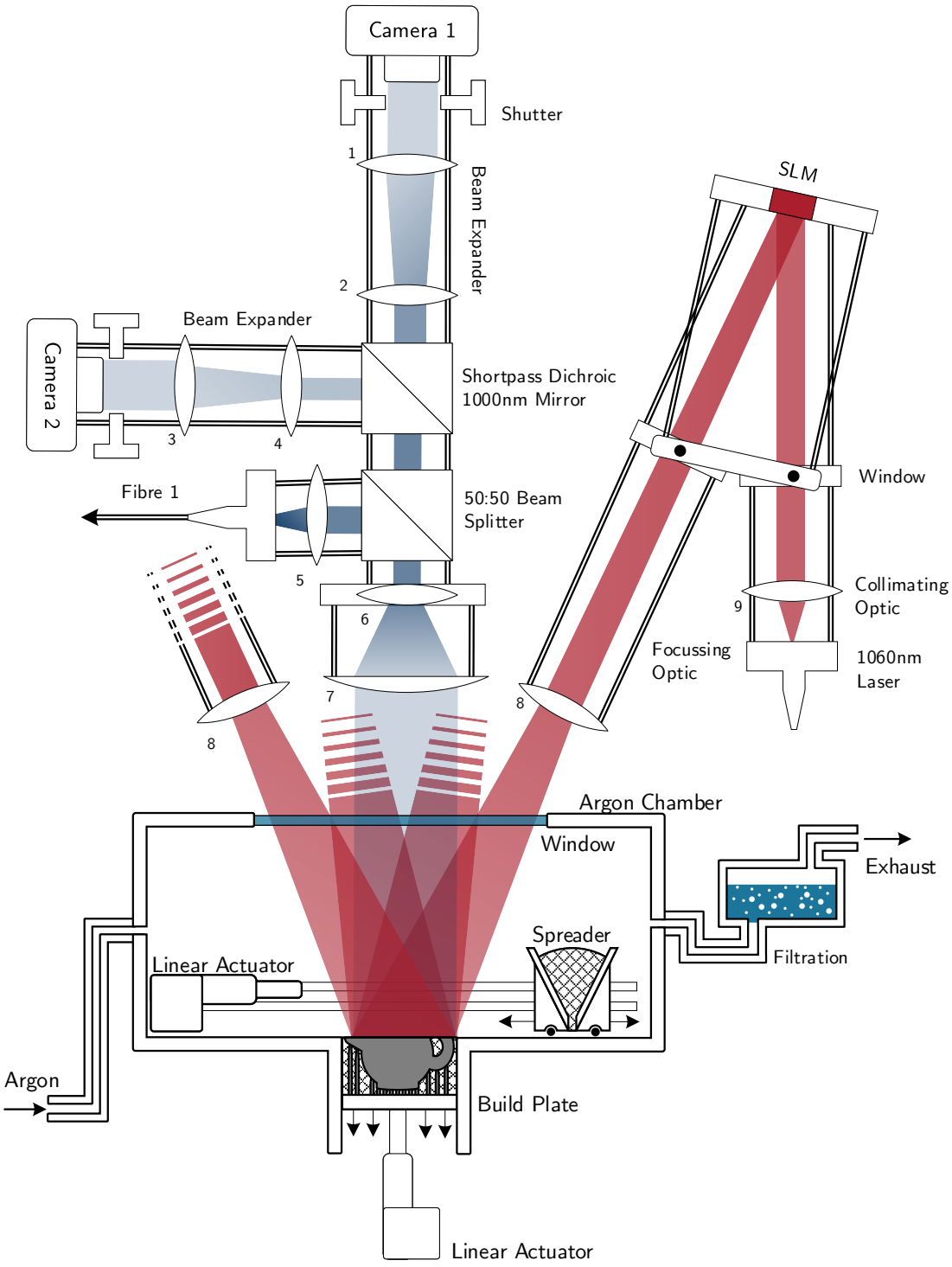


Fig. 6.10 Demonstrator schematic

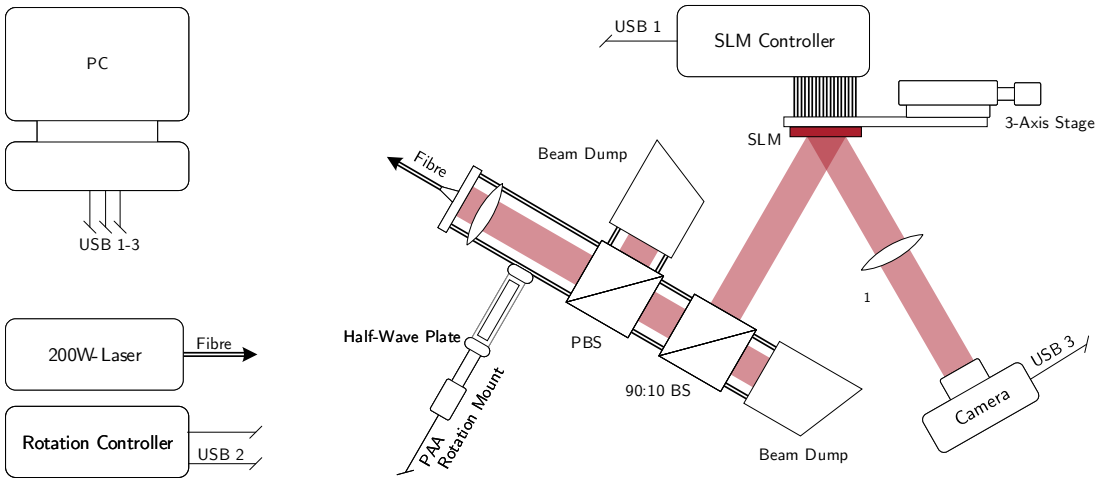


Fig. 6.11 High-power test rig schematic

6.3.3 Progress To Date

While much work remains to be done, some progress has been made.

Components

A number of components have been acquired including:

- **SLM:** The SLM has been chosen as a Santec SLM-300, Table 6.4. Released during the second year of the PhD, this is the first commercial SLM available with a rated power of over 300W CW.
- **Processing laser:** Due to availability, the processing laser has been selected as an SPI SP-200C-0001, a 1090nm, 200W fibre laser. More information is given in Table 6.5
- **Visible camera:** Selected as a ThorLabs Kiralux, 12.3 megapixel monochrome CMOS camera. This will allow for imaging of the bed at resolutions greater than the resolution of the hologram projected.
- **Thermographic camera:** Due to availability, the camera for imaging at 8-12µm wavelengths will be both a Thermoteknix Miracle 307K and Testo 868. These will be used for confirmation of the in-line process model.

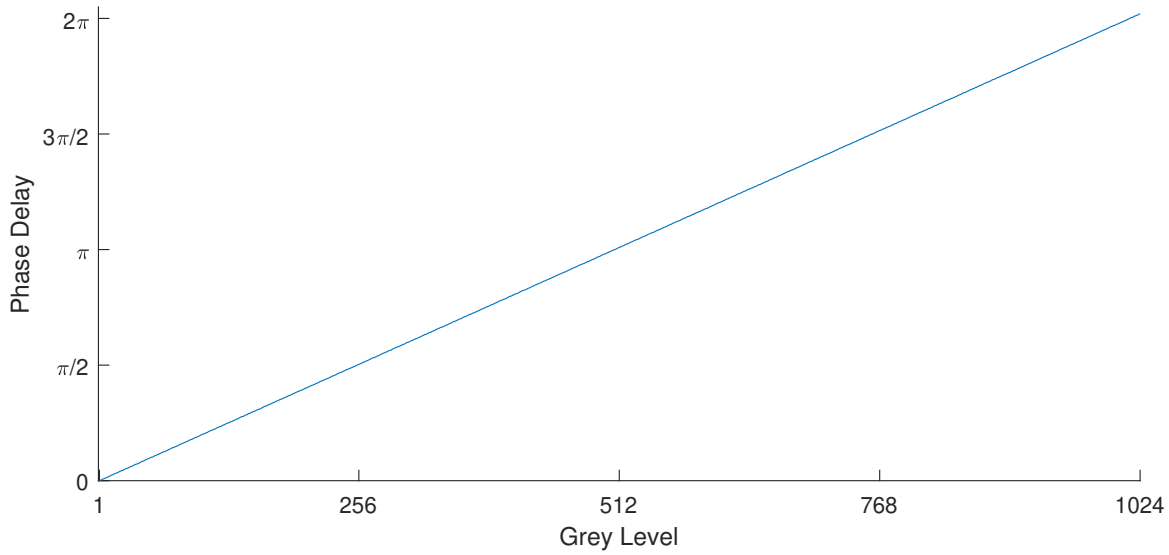


Fig. 6.12 SLM-300 calibration

These four elements together form the bulk of the cost of the project and more than meet the requirements.

Performance Test

In order to characterise the SLM-300 at high power a calibration was required. The system used for this was a simple arrangement with schematic shown in Figure 6.11. The rendered and achieved versions are shown in Figures 6.13 and 6.14 respectively.

The SLM-300 worked well with a phase response curve as shown in Figure 6.12 and a total measured phase inaccuracy of less than 0.001π. This was done using the method of showing a blazed grating with varied phase level and measuring the total power recorded in the first diffraction order.

Replay drift measurements also observed significantly less than a pixel of movement in the far field over the frame cycle of the SLM in contrast to the behaviour of the Jasper JD8714 used in Section 6.2.4.

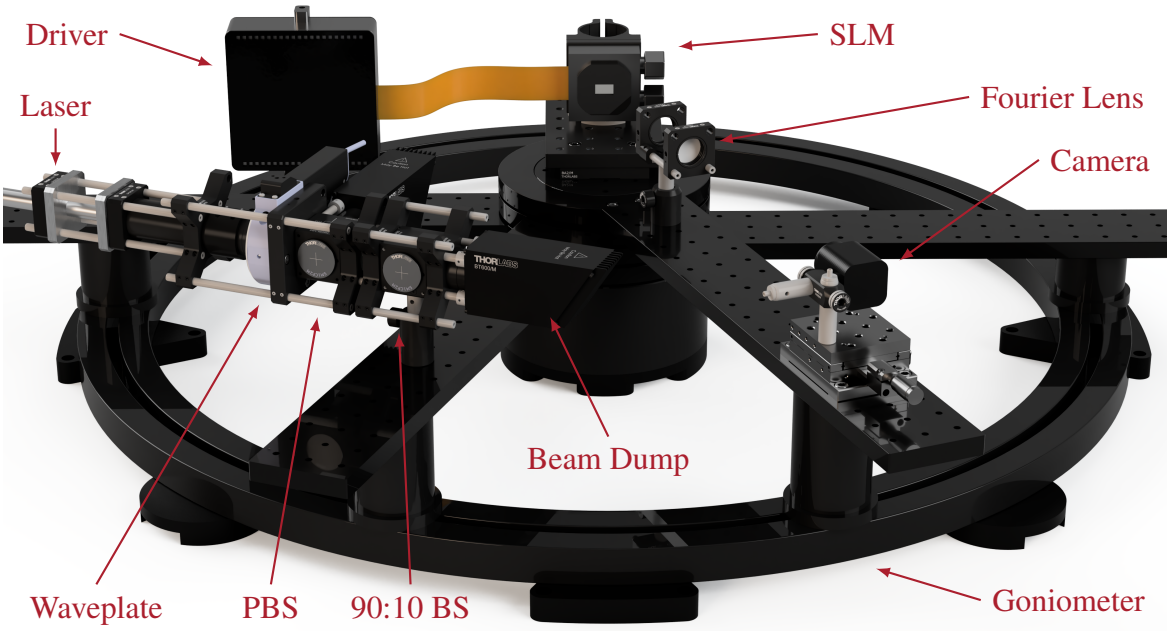


Fig. 6.13 High-power test rig render

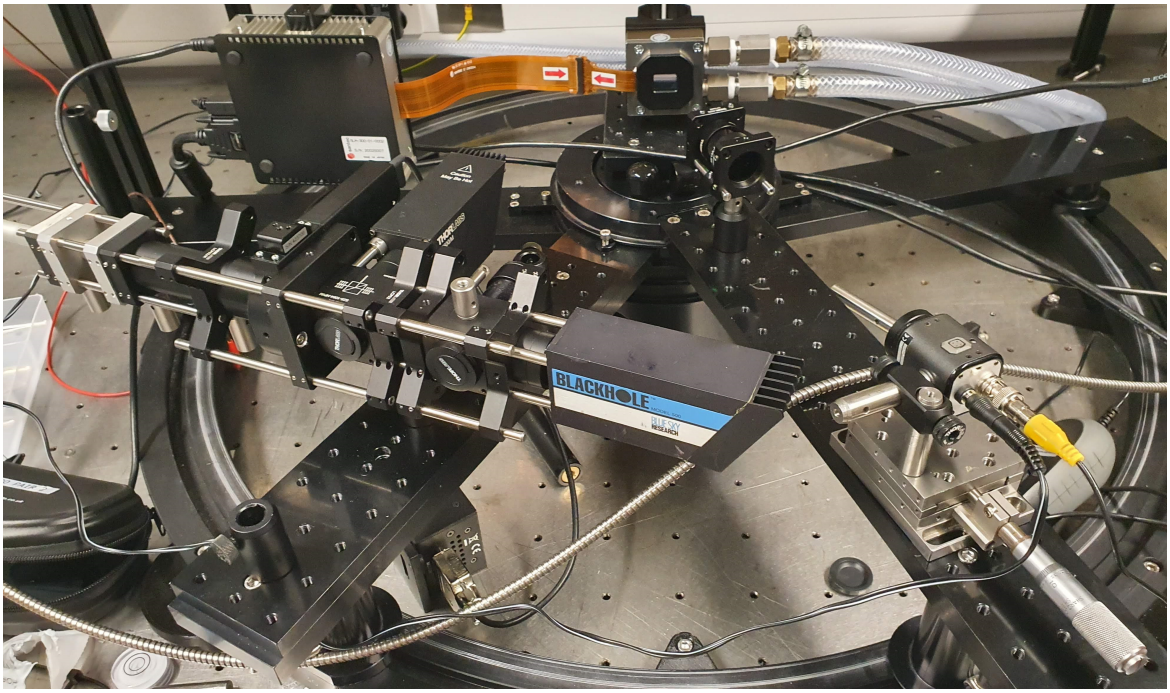


Fig. 6.14 High-power test rig photograph

Photogrammetric Alignment

One challenge of the metal process is the thermal deformations expected in any machine. In order to map and characterise these as well as record the progress of layer exposure, a registration mark scheme using fiducial tags has been designed. Currently confirmed only by simulation, the scheme used in Figure 6.15 has been followed. The camera is aligned relative to the bed via a set of fixed registration marks using the AliceVision cctag library [288]. An additional set of registration marks are then projected onto the bed by the holographic projector and used to determine the location of the projector.

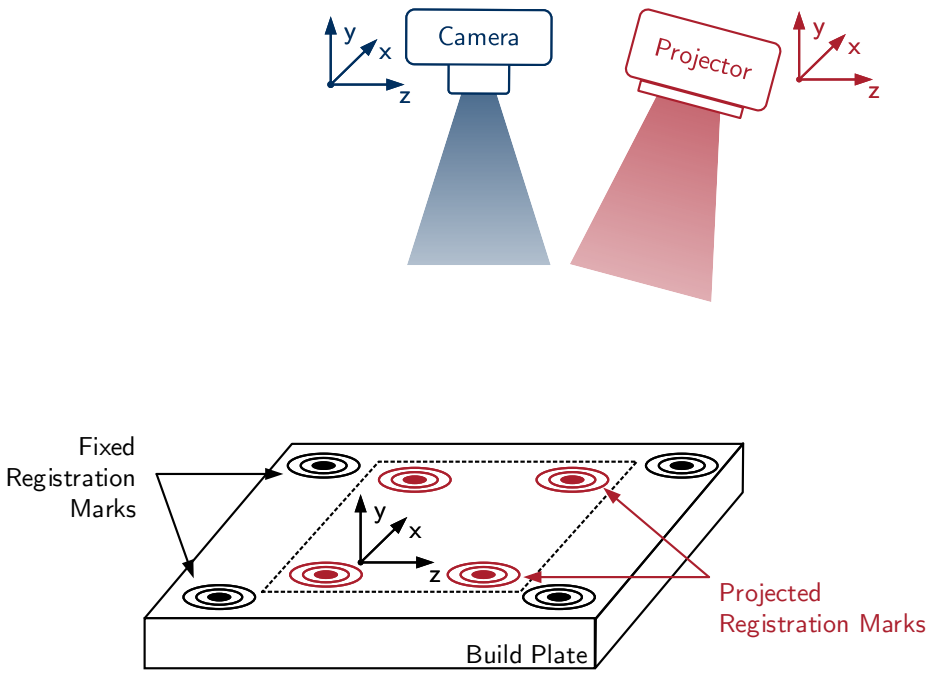


Fig. 6.15 Alignment registration marks

The key advantage of this system is that it can be defined afocally as the impulse rings used in cctags are robust when perturbed from focus due to their relation ship to the Bessel patterns used on the SLM to excite them.

Thermal Model

A finite difference model for 2D and 3D thermal analysis of the print has been developed. Similar to the photogrammetric system, this has been demonstrated theoretically but has yet to be verified against a real-world case.

6.3.4 Future Work

In addition to completing the work started in the previous section, a number of additional elements will be required.

Material Choice

The first piece of work to be done during the postdoc is in the choice of powder. For demonstration, a low coefficient of thermal expansion (CTE) to stiffness ratio is required. This rules out titanium in particular but also a number of alloys. Additionally, the overall heat input required should be low to reduce exposure times. This suggests materials with lower specific heat capacities. The materials Inconel In625-0402 and In718-0405 are currently favourites for operation but a final decision is being delayed while AM partnership options are being explored.

Slicing

In order to convert a 3D geometry into holograms, a slicing algorithm is required. The resin printer, Section 6.2, used Autodesk's Netfabb application which was sufficient for that application. As the hologram projection will use knowledge of the feature relationships between slices as well as within a slice, a custom slicer will be used to account for this. One challenge here will be the fact that many slicers depend on the 'slic3r' package which is released under the GPL and is therefore not suitable for a potentially patentable approach [289].

Chamber

The key requirement on the chamber is that it contain the powder, oxidised material, gas and laser emission while maintaining a consistent temperature. A planned future project investigating patterning also requires a low vacuum. As a result, the chamber will be designed out of stainless 304 steel due to its high thermal conductivity and lower outgassing than competing materials such as aluminium[290].

The top of the chamber will have three wedged vacuum windows, one for the projector, one for the vision system and one for monitoring. Bolted flanges with vacuum seals and copper gaskets will be used for the access ports and assembly.

Optics

The projector will follow a similar pattern to the system used for the resin printer in Section 6.2, albeit with higher-power optics and added water cooling. The camera will use a notch filter to remove the 1090nm laser emission scattered from the powder while still allowing imaging in the visible and IR regions. The optical system will be thermally isolated from the chamber to reduce thermal distortions and the photogrammetric system used to keep the alignment calibrated.

Characterisation

Unlike in the resin case where an assessment of form was sufficient, metal powder printing will require an array of different tests to characterise material properties including roughness, residual stress, porosity, voids, fatigue and alloy loss. Available processes are discussed in greater detail in Chapter 3.

Commercialisation and Partners

The final element to be completed will be commercialisation. As this process is likely to offer significant benefits to commercial AM users and manufacturers, it is expected that interest will be high. The IP protection process has been started with Cambridge Enterprise and discussions started with potential partners.

6.3.5 Summary

This section has introduced the work done so far on the powder demonstrator being built as part of an EPSRC funded project investigation holographic additive manufacture. While much preparatory work has been completed, much still remains and will be carried out over the coming months as part of a postdoctoral research position.

6.4 Conclusion

This chapter has introduced two demonstrator systems developed for this thesis. The first, Section 6.2, was a full areal printer in resins using a process termed (Holographic Additive Manufacture). This demonstrator was completed and the design, operation, performance and results were discussed. The second remains a work in progress and is targeted at operation with plastic and metal powders. Key design criteria and expected work packages were also introduced. In the final chapter, we shall discuss the broader results of this project as well as future work and expected research directions.

Chapter 7

Conclusion

Chapter 7

Conclusion

This thesis has presented the work carried out over the last three years of the PhD project. This chapter briefly summarises the key findings in Section 7.1 before quickly summarising the included appendices in Section 7.2. Section 7.3 then proceeds to discuss possible future work.

7.1 Summary and Key Findings

The fields of additive manufacture and holography were introduced in Chapter 1. The mathematics of holography and additive manufacture were expanded on in Chapters 2 and 3 respectively.

Chapter 4 then introduced the first deliverable for the project, an open-source software package for hologram generation with state-of-the-art performance on Nvidia GPUs as well as some of the batch processing, region masking, image manipulation and parameter framework features included.

Five new families of algorithms invented during the PhD were introduced in Chapter 4.3 as well as the experimental configuration used for testing and their applications and relevance to the AM project. These were then expanded on in the following five sections.

- **Section 5.2** introduced sorted pixel selection (SPS) which offered a novel on search algorithms including direct search and simulated annealing [266].
- **Section 5.3** introduced the single-transform time-multiplexed (STTM) algorithm which offered order of magnitude improvements in computational runtimes when compared to previous state-of-the-art time-multiplexed algorithms such as OSPR [267].
- **Section 5.4** introduced the holographic predictive search (HPS) family of algorithms that offered multiple orders of magnitude improvements in runtime over traditional search techniques at the expense of a more complex mathematical formulation [268, 269].
- **Section 5.5** introduced sympathetic quantisation (SQ) which offered significant improvements in visual quality in single iteration systems such as those used in holographic headsets. [270].
- **Section 5.6** introduced the linear time (LT) algorithm which offered a breakthrough linear time performance for phase sensitive replay fields as well as guaranteed global minimum performance [271].

This was followed by the introduction of a demonstrator for holographic additive manufacture in Section 6.2. This was a stereolithography using holographic illumination process that matched the performance of commercially available LCD based stereolithography machines.

Finally Section 6.3 introduced the ongoing work into a second demonstrator at high power for metallic powder printing and outlines the expected next step as part of an EPSRC funded grant.

7.2 Appendices

A selection of the side projects carried out during the PhD are summarised in the appendices.

- **Appendix A** investigates the computational performance of the OSPR algorithm introduced in Section 2.4.5
- **Appendix B** presents a heuristic analysis of the performance factors effecting the runtimes of iterative algorithms such as Gerchberg-Saxton. This also serves as a test-bed for many of HoloGen's more advanced features.

- **Appendix C** summarises a published journal paper that demonstrates the relative limitations of multi-amplitude devices and presents heuristic and mathematical evidence for the reasoning behind this.
- **Appendix D** summarises one aspect of a side project looking at embedding the hologram generation process into an FPGA, the use of look-up tables for the phase randomisation step.
- **Appendix E** summarises an experimental investigation of the damage thresholds of ferroelectric SLMs.
- **Appendix F** supports Appendix E in summarising a method for finding the ellipticity of an unknown polarisation state in an automated system.
- **Appendix G** presents a detailed overview of the individual libraries and key classes making up the HoloGen application.
- **Appendix H** gives a brief summary of the individual libraries used in the ResinPrinter application powering the resin printer discussed in Section 6.2
- **Appendix I** lists available additional resources including source code, 3D models, simulations and other resources that may be of use to future readers.

7.3 Future Work

While this thesis has covered a lot of ground and successfully demonstrated that holographic additive manufacture is both possible and practical, more work remains to be done for scaling this to the targeted high-power cases of powder based AM. The work to date was summarised in Section 6.3. This also summarised the work-packages to be carried out over the next year as part of an EPSRC funded post-doctoral position.

Expected components of this include:

- **Mechanical machine** - As part of demonstrating principle, a metallic powder printer has been designed. This still needs to be constructed, tested, calibrated and validated.
- **Mathematical model** - A finite-difference package has been constructed for measuring the relative performance of holographic AM with conventional L-PBF. This was

mentioned in Section 6.3 but real-world measurement data will be required for validating its performance.

- **Real-time compensation** - As yet unpublished, significant work has been done for photogrammetry, fiducial alignment and thermal imaging of AM processes in real-time. As before, real-world measurement data will be required for validating its performance.
- **Test parts and validation** - The final stage will include the printing of common test parts for validation of the process.

As well as continuing work on the high-power additive manufacture, a number of ongoing side projects require publication.

- **Holographic optical fibre modes** - A collaboration with Mr Ralf Mouthaan into novel generation algorithms for fibre modes.
- **Quantum computer hologram generation** - A collaboration with Dr George Gordon using a quantum computer for generating holograms.
- **Focal depth tracking for improved 3D holographic reconstruction** - A collaboration with Mr Jamie Lake and Dr Hannah Joyce using eye-tracking for tuning the quality of holograms with focal depth.
- **Holographic multi-beam interference lithography** - A collaboration with Dr Hannah Joyce, Dr Jack Alexander-Webber and Dr Calum Williams proposing to use holographic beam shaping as a component step in multi-beam interference lithography.
- **Hybrid printing for custom microfluidic devices** - A collaboration with Mr Elliot Medcalf and Professor Florian Hollfelder exploring the opportunities for a hybrid 3D printing approach for printing microfluidic devices.

The first two of these are substantially complete but require additional work in writing up while the remaining three are ongoing.

7.4 Conclusion

This thesis has covered an initial demonstration of a new process in additive manufacture, known as holographic additive manufacture or HAM. Two main areas can be identified in

this work. The first covers the necessary software and algorithms, the second the design and characterisation of an initial resin demonstration system and designs for an upcoming follow-on powder demonstrator.

7.4.1 Software and Algorithms

The HoloGen package introduced in Chapter 4 introduced the world's first open-source hologram generation suite targeting deployment on graphics cards and the full gambit of different algorithms and device types. While individual algorithms have been deployed to graphics cards before, this is the first time that it has been brought together in one place along with detailed guides on the selection criteria for different techniques. Initial interest has been high, and as commercial investment in holography grows, the number of users of HoloGen is also expected to grow. As computer-generated holography moves from selected niche applications towards being a cornerstone of commercial applications in imaging and projection, it is hoped that HoloGen will become an internationally recognised platform for teaching, understanding, deploying and application of computer-generated holograms.

HoloGen also serves as a necessary foundation for holographic additive manufacture (HAM). While real-time holography implementations are available, they are targeted at visual applications without the generality of HoloGen. Importantly, while the demonstrator used here only incorporates one SLM, HoloGen is the first reported platform designed for the generation of multiple synchronous frames for use with multiple overlapping projections. It is also the first to include a 3D geometry slicer package as a source of the images and the first to be use captured imaging data to adjust holograms further down the pipeline. Future work on HoloGen is discussed in Appendix G.7 and will focus on extending the range of devices and platforms on which HoloGen can be deployed.

While HoloGen provides state-of-the-art performance for existing algorithms, Chapter 4.3 introduced some of the novel algorithms developed for the HAM process. Holograms for HAM are split into two classes, those that are pre-generated and those generated in real-time. Pre-generated holograms offer improved image quality at the expense of high memory use, slower runtimes and an inability to incorporate feedback from process monitoring.

The first pre-generated holography algorithm, sorted pixel selection (SPS), introduced targeted pre-generated holograms on a binary SLM. While the eventual demonstrator incorporated a multi-phase device, SPS offers performance improvements of up to 16.5%

over common algorithms such as direct search (DS) with extremely low overhead and can be incorporated in the majority of binary hologram generation algorithms. The excellent cost-benefit ratio of SPS means that it is an excellent improvement in search algorithms where simplicity is critical.

While SPS was used in early version of the demonstrator, the process was soon adapted to use holographic predictive search (HPS). Also primarily targeted at binary SLMs, this offered performance improvements of $50\times$ in convergent error and convergence time compared to existing search algorithms at the expense of additional mathematical complexity and reduced flexibility. The common categories found in binary SLM configurations include: phase vs amplitude modulation; real or complex target field; and Fraunhofer vs Fresnel diffraction. HPS offers significant performance improvements over the previous state-of-the-art in every case with lower convergent mean squared error and greater structural-similarity as well as improved runtimes. This comes at the expense of increased mathematical complexity which is discussed in Section 5.4.

The true power of HAM comes to the fore, however, in real-time applications where the exposure pattern can be adapted in-line with vision system feedback. This necessitates a real-time algorithm. While multiplexed approaches such as one-step phase-retrieval (OSPR) are available, this thesis proposed the single-transform time-multiplexed (STTM) algorithm, an alternative approach with better convergent error and an 80% reduction in generation time. Like SPS and HPS, STTM targets binary SLMs. The real-time constraint, however, means that the time-averaging of 1000s of low-quality frames per second produces a more accurate result than fewer, better optimised frames. At the time of writing, STTM offered the best performance to runtime ratio of any reported algorithm targeting high-frame-rate binary devices.

During the early stage of work on the PhD it was intended to operate with an array of micro projectors based on binary ferroelectric SLMs. Due to delays in the development of the micro-projectors, the version of the demonstrator discussed in Chapter 6 used multi-phase SLMs. These have much lower frame-rates than ferroelectric LCs but offer more phase control. Pre-generated frames for this were produced using the widely known Gerchberg-Saxton algorithm. A new algorithm was introduced, however, for real-time frames known as sympathetic quantisation (SQ). This used properties of the hologram generation process to produce much better quality initial frames than achievable using GS at the expense of limited ability for further improvement using iteration. With restricted computing power available,

this allowed for real-time operation of a 4K SLM at medium frame-rates without the use of high-end graphics cards.

The only algorithm introduced in Chapter 4.3 that was not directly used during development of the AM demonstrator in this project was the linear time (LT) algorithm. This was originally conceived for fibres with the end goal that complex control of the phase input to a fibre would allow control of the output. It was originally intended that this could be then incorporated into a more traditional extrusion or jetting AM progress that would allow for customised profile delivery without the use of galvoscanners. The technique was never developed due to Covid restrictions but is included here due to its greatly improved generation time performance for phase applications of up to $50,000\times$.

7.4.2 Demonstrator

While the software and algorithms provide a necessary foundation for the HAM process, technique demonstration is required to ensure that real-world implementation of theoretical processes is achievable. To leverage this, a HAM demonstrator targeting printing in resins was introduced in Section 6.2. The goal of this was to demonstrate that the HAM process could match commercial systems in speed, print quality and print flexibility for existing material processes. A well regarded commercial resin stereolithography (SLA) machine was chosen as a benchmark, the Prusa SL1. By showing a sequence of holographically shaped images onto a liquid resin, the resin demonstrator was able to produce cured resin objects of a quality matching that of the commercial machine with the inclusion of real-time monitoring. HAM is unlikely to replace SLA in this field due to its increased complexity and cost but this offers a good benchmark for HAM performance.

The primary motivation of HAM, however, is its power handling and greater resolution flexibility. While SLA is tied to the resolution of the display it uses, HAM is not and is capable of greatly reduced feature sizes without compromising on build area. Additionally, SLA cannot scale to high power, again due to the liquid crystal displays. By contrast, commercial SLMs have recently become available that can handle 100s of watts of power, sufficient in practice to melt areas of plastic or metal powders. Prior to the Covid-19 pandemic it had been intended to demonstrate both of these benefits with the design and implementation of a powder demonstrator for the HAM process before the end of the PhD. Unlike the SLA system, this would have incorporated a 200W IR laser and appropriate high power capacity SLM. Inevitable delays due to building and supplier closure as well as lab restriction and

technician availability meant that that this second demonstrator has been delayed until after submission of the PhD thesis.

7.4.3 Going Forward

While this PhD has introduced many of the tools required for using the HAM process as well as the first process demonstration, much work remains to be done in-order to increase its impact and to transition it from the lab to industrial use.

First on the list of necessary work is the powder demonstrator originally planned for completion during the PhD. Progress on this is good and initial indications are promising that this will provide conclusive evidence that HAM offers greatly improved powder printing rates when compared to the current state-of-the-art while also offering improved thermal performance and material behaviours. This research is ongoing, however, and is supported by a two year post-doctoral grant. Once a powder demonstrator is completed, work will commence on exploiting commercialisation avenues. To ensure maximum industry take-up, this is expected to be primarily via licensing as opposed to a commercial spin-out.

In parallel to exploring the high-power capabilities of HAM, it is intended to explore the promise of decreased feature sizes and resolution. If successful, commercialisation avenues for this will also be explored.

In conclusion after introducing the fields of holography and additive manufacture this thesis has sought to investigate high-rate additive manufacture using holographic beam shaping. To enable this, a new open-source package for hologram generation with state-of-the-art performance has been built. Five new families of algorithms for hologram generation, often with multiple orders of magnitude performance improvements, have been introduced. A holographic resin 3D printer demonstrator has been characterised and tested. Finally a detailed plan of future work has been presented.

References

References

[1] T. DebRoy, H. L. Wei, J. S. Zuback, T. Mukherjee, J. W. Elmer, J. O. Milewski, A. M. Beese, A. Wilson-Heid, A. De, and W. Zhang, "Additive manufacturing of metallic components – Process, structure and properties," *Progress in Materials Science*, vol. 92, pp. 112–224, 2018.

[2] ASTM F2792 - 09, *Standard Terminology for Additive Manufacturing Technologies*. ASTM International, 2009.

[3] T. T. Wohlers, *Wohlers Report 2019: 3d printing and additive manufacturing state of the industry. Terry Wohlers*. 2019.

[4] V. Bhavar, P. Kattire, P. Pawar, S. Khot, K. Gujar, and R. Singh, "A Review on Powder Bed Fusion Technology of Additive Manufacturing," in *Processes, Applications, and Performance of Materials in Additive Manufacturing*, pp. 1–2, 2014.

[5] C. K. Chua, K. F. Leong, and C. S. Lim, *Rapid prototyping: principles and applications*. World Scientific, 2010.

[6] J. Xiong, Y. Lei, H. Chen, and G. Zhang, "Fabrication of inclined thin-walled parts in multi-layer single-pass GMAW-based additive manufacturing with flat position deposition," *Journal of Materials Processing Technology*, vol. 240, pp. 397–403, 2017.

[7] C. G. Pickin, S. W. Williams, and M. Lunt, "Characterisation of the cold metal transfer (CMT) process and its application for low dilution cladding," *Journal of Materials Processing Technology*, vol. 211, no. 3, pp. 496–502, 2011.

[8] M. R. Johnson, "The Galileo High Antenna Deployment Anomaly," *The 28th Aerospace Mechanisms Symposium*, vol. 28, pp. 359–377, 1994.

[9] D. Ding, Z. Pan, D. Cuiuri, and H. Li, "Wire-feed additive manufacturing of metal components: technologies, developments and future interests," *International Journal of Advanced Manufacturing Technology*, vol. 81, no. 1-4, pp. 465–481, 2015.

[10] S. Kalpakjian, S. R. S. Schmid, and H. Musa, *Manufacturing Engineering and Technology*. Pearson Upper Saddle River, NJ, USA, 2009.

[11] D. E. Schick, R. M. Hahnlen, R. Dehoff, P. Collins, S. S. Babu, M. J. Dapino, and J. C. Lippold, "Microstructural characterization of bonding interfaces in Aluminum 3003 blocks fabricated by ultrasonic additive manufacturing," *Welding Journal*, vol. 89, no. 5, pp. 105–115, 2010.

[12] J. L. de Bougrenet de la Tocnaye and L. Dupont, "Complex amplitude modulation by use of liquid-crystal spatial light modulators," *Applied Optics*, vol. 36, no. 8, p. 1730, 1997.

[13] N. Collings, "Advanced spatial light modulator technology," in *Communications and Photonics Conference (ACP), 2012 Asia*, pp. 1–2, IEEE, 2012.

[14] T. D. Wilkinson, *Fundamentals of Optics For Display - Optical Modulation*, pp. 48–67. 2012.

[15] P. Raynes, *Twisted Nematic and Supertwisted Nematic LCDs*, pp. 2077–2090. Cham: Springer International Publishing, 2016.

[16] L. J. Hornbeck, "Active yoke hidden hinge digital micromirror device," July 1996. US Patent 5,535,047.

[17] M. R. Douglass, "Lifetime estimates and unique failure mechanisms of the digital micromirror device (dmd)," in *Reliability Physics Symposium Proceedings, 1998. 36th Annual. 1998 IEEE International*, pp. 9–16, IEEE, 1998.

[18] D. B. Beasley, M. W. Bender, J. Crosby, T. Messer, and D. A. Saylor, "Advancements in the micromirror array projector technology," in *Technologies for Synthetic Environments: Hardware-in-the-Loop Testing VIII*, vol. 5092, pp. 71–83, International Society for Optics and Photonics, 2003.

[19] D. B. Beasley, M. Bender, J. Crosby, S. McCall, T. Messer, and D. A. Saylor, "Advancements in the micromirror array projector technology ii," in *Technologies for Synthetic Environments: Hardware-in-the-Loop Testing X*, vol. 5785, pp. 68–80, International Society for Optics and Photonics, 2005.

[20] D. M. Bloom, "Grating light valve: revolutionizing display technology," in *Projection Displays III*, vol. 3013, pp. 165–172, International Society for Optics and Photonics, 1997.

[21] P. W. M. Tsang and T. C. Poon, "Review on the State-of-the-Art Technologies for Acquisition and Display of Digital Holograms," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 886–901, 2016.

[22] Y. Huang, E. Liao, R. Chen, and S.-T. Wu, "Liquid-Crystal-on-Silicon for Augmented Reality Displays," *Applied Sciences*, vol. 8, no. 12, pp. 1–17, 2018.

[23] C. Bay, N. Hübner, J. Freeman, and T. Wilkinson, "Maskless photolithography via holographic optical projection," *Opt. Lett.*, vol. 35, pp. 2230–2232, Jul 2010.

[24] T. D. Wilkinson, C. Bay, and J. Freeman, "Maskless photolithography by holographic optical projection," in *Imaging and Applied Optics 2014*, p. DW1B.3, Optical Society of America, 2014.

[25] J. P. Parry, R. J. Beck, J. D. Shephard, and D. P. Hand, "Application of a liquid crystal spatial light modulator to laser marking," *Applied Optics*, vol. 50, no. 12, pp. 1779–1785, 2011.

[26] P. Bartolo and J. Gaspar, "Metal filled resin for stereolithography metal part," *CIRP annals*, vol. 57, no. 1, pp. 235–238, 2008.

[27] H. X. Nguyen, H. Suen, B. Poudel, P. Kwon, and H. Chung, "Development of an innovative, high speed, large-scaled, and affordable metal additive manufacturing process," *CIRP Annals*, 2020.

[28] G. Račiukaitis, E. Stankevičius, P. Gečys, M. Gedvilas, C. Bischoff, E. Jäger, U. Umhofer, and F. Völklein, "Laser processing by using diffractive optical laser beam shaping technique.," *Journal of Laser Micro / Nanoengineering*, vol. 6, no. 1, 2011.

[29] F. M. Dickey, L. S. Weichman, and R. N. Shagam, "Laser beam shaping techniques," in *High-Power Laser Ablation III*, vol. 4065, pp. 338–349, International Society for Optics and Photonics, 2000.

[30] A. Okunkova, M. Volosova, P. Peretyagin, Y. Vladimirov, I. Zhirnov, and A. Gusarov, "Experimental approbation of selective laser melting of powders by the use of

non-gaussian power density distributions," *Physics Procedia*, vol. 56, pp. 48–57, 2014.

[31] Y. Li, "Light beams with flat-topped profiles," *Optics Letters*, vol. 27, no. 12, pp. 1007–1009, 2002.

[32] J. A. Hoffnagle and C. M. Jefferson, "Design and performance of a refractive optical system that converts a gaussian to a flattop beam," *Applied Optics*, vol. 39, no. 30, pp. 5488–5499, 2000.

[33] S. Zhang, G. Neil, and M. Shinn, "Single-element laser beam shaper for uniform flat-top profiles," *Optics Express*, vol. 11, no. 16, pp. 1942–1948, 2003.

[34] K. Pangovski, M. Sparkes, A. Cockburn, W. O'Neill, P. S. Teh, D. Lin, and D. Richardson, "Control of material transport through pulse shape manipulation—a development toward designer pulses," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 5, pp. 51–63, 2014.

[35] J. W. Goodman, *Introduction to Fourier Optics*. Roberts & Co. Publishers, 3rd ed., 2005.

[36] E. Hecht, *Optics*. Pearson Education, 4th ed., 2017.

[37] D. J. Griffiths, *Introduction to Electrodynamics*. Cambridge University Press, 4 ed., 2017.

[38] G. Makey, Ö. Yavuz, D. K. Kesim, A. Turnalı, P. Elahi, S. Ilday, O. Tokel, and F. Ö. Ilday, "Breaking crosstalk limits to dynamic holography using orthogonality of high-dimensional random vectors," *Nature Photonics*, vol. 13, no. 4, p. 251, 2019.

[39] J. W. Goodman, *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2007.

[40] J. Freeman, *Visor Projected Helmet Mounted Display for Fast Jet Aviators using a Fourier Video Projector*. PhD thesis, University of Cambridge, 2009.

[41] J. P. Freeman, T. D. Wilkinson, and P. Wisely, "Visor projected HMD for fast jets using a holographic video projector," in *Proc. SPIE*, vol. 7690, p. 76901H, 2010.

[42] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[43] A. J. Cable, E. Buckley, P. Mash, N. A. Lawrence, T. D. Wilkinson, and W. A. Crossland, "Real-time Binary Hologram Generation for High-quality Video Projection Applications," in *SID Symposium Digest of Technical Papers*, vol. 35, pp. 1431–1433, 2004.

[44] V. Lakshminarayanan and A. Fleck, "Zernike polynomials: A guide," *Journal of Modern Optics*, vol. 58, p. 1678, 2011.

[45] D. Brown, "Decentering Distortion of Lenses," *Photometric Engineering*, vol. 32, no. 3, pp. 444–462, 1966.

[46] J. C. Wyant and K. Creath, "Basic wavefront aberration theory for optical metrology," *Applied Optics and Optical Engineering*, vol. Xl, no. part 2, pp. 1–53, 1992.

[47] H. Gross, F. Blechinger, and B. Achtner, "Handbook of Optical Systems," *Handbook of Optical Systems*, vol. 4, pp. 1–1064, 2015.

[48] Z. Wang, a. C. Bovik, H. R. Sheikh, and E. P. Simmoncelli, "Image quality assessment: form error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612.

[49] USC-SIPI Database, "www.sipi.usc.edu/database/," accessed on: 2020-12-07.

[50] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[51] J. R. Fienup and C. C. Wackerman, "Phase-retrieval stagnation problems and solutions," *Journal of the Optical Society of America A*, vol. 3, no. 11, p. 1897, 1986.

[52] V. Kettunen, "Review of iterative Fourier-transform algorithms for beam shaping applications," *Optical Engineering*, vol. 43, no. 11, p. 2549, 2004.

[53] J. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," in *Applications of Digital Image Processing III*, vol. 207, pp. 2–14, International Society for Optics and Photonics, 1979.

[54] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Optics Letters*, vol. 3, no. 1, p. 27, 1978.

[55] A. Jesacher, A. Schwaighofer, S. Fürhapter, C. Maurer, S. Bernet, and M. Ritsch-Marte, "Wavefront correction of spatial light modulators using an optical vortex image," *Optics Express*, vol. 15, no. 9, pp. 5801–5808, 2007.

[56] J. Bengtsson, "Kinoform design with an optimal-rotation-angle method," *Applied Optics*, vol. 33, no. 29, pp. 6879–6884, 1994.

[57] M. Johansson and J. Bengtsson, "Robust design method for highly efficient beam-shaping diffractive optical elements using an iterative-Fourier-transform algorithm with soft operations," *Journal of Modern Optics*, vol. 47, pp. 1385–1398, July 2000.

[58] F. Wyrowski, "Iterative quantization of digital amplitude holograms," *Applied Optics*, vol. 28, no. 18, p. 3864, 1989.

[59] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, p. 2758, 1982.

[60] J. S. Liu and M. R. Taghizadeh, "Iterative algorithm for the design of diffractive phase elements for laser beam shaping.," *Optics Letters*, vol. 27, no. 16, pp. 1463–1465, 2002.

[61] B. Z. Dong, Y. Zhang, B. Y. Gu, and G. Z. Yang, "Numerical investigation of phase retrieval in a fractional Fourier transform," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 14, no. 10, pp. 2709–2714, 1997.

[62] Z. Liu, L. Xu, Q. Guo, C. Lin, and S. Liu, "Image watermarking by using phase retrieval algorithm in gyrator transform domain," *Optics Communications*, vol. 283, no. 24, pp. 4923–4927, 2010.

[63] J. S. Liu, M. J. Thomson, and M. R. Taghizadeh, "Automatic symmetrical iterative Fourier-transform algorithm for the design of diffractive optical elements for highly precise laser beam shaping," *Journal of Modern Optics*, vol. 53, no. 4, pp. 461–471, 2006.

[64] B. K. Jennison, J. P. Allebach, and D. W. Sweeney, "Direct binary search computer-generated holograms: an accelerated design technique and measurement of wavefront quality," in *SPIE - O-E LASE '89: Optoelectronics and Laser Applications in Science and Engineering*, pp. 2–9, International Society for Optics and Photonics, 1989.

[65] K. Brian, K. Jennison, and P. Jan, "Iterative approaches to computer-generated holography," *Optical Engineering*, vol. V, no. 6, pp. 629–637, 2015.

[66] B. B. Chhetri, S. Yang, and T. Shimomura, "Stochastic Approach in the Efficient Design of the Direct-Binary-Search Algorithm for Hologram Synthesis," *Applied Optics*, vol. 39, no. 32, p. 5956, 2000.

[67] V. Boutenko and R. Chevallier, "Second order direct binary search algorithm for the synthesis of computer-generated holograms," *Optics Communications*, vol. 125, no. 1-3, pp. 43–47, 1996.

[68] A. Palani, *Development of an optical system for dynamic evaluation of phase recovery algorithms*. Phd thesis, University of Cambridge, 2014.

[69] F. Wyrowski, "Diffractive optical elements: iterative calculation of quantized, blazed phase structures," *Journal of the Optical Society of America A*, vol. 7, no. 6, p. 961, 1990.

[70] A. Kirk and T. J. Hall, "Design of binary computer generated holograms by simulated annealing: coding density and reconstruction error," *Optics Communications*, vol. 94, no. 6, pp. 491–496, 1992.

[71] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *et al.*, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[72] C.-S. Kim, "Simple distortion-invariant optical identification tag based on encrypted binary-phase computer-generated hologram for real-time vehicle identification and verification," *Optical Engineering*, vol. 49, no. 11, p. 115801, 2010.

[73] A. J. Cable, *Real-time high-quality two and three-dimensional holographic video projection using the one-step phase retreival approach*. Phd thesis, Cambridge University, 2007.

[74] D. H. Kelly and D. van Norren, "Two-band model of heterochromatic flicker," *Journal of the Optical Society of America*, vol. 67, no. 8, pp. 1081–1091, 1977.

[75] Y. Peng, S. Choi, N. Padmanaban, J. Kim, and G. Wetzstein, "Neural holography," in *ACM SIGGRAPH 2020 Emerging Technologies*, pp. 1–2, 2020.

[76] N. Yoshikawa, M. Itoh, and T. Yatagai, "Quantized phase optimization of two-dimensional Fourier kinoforms by a genetic algorithm.," *Optics Letters*, vol. 20, no. 7, pp. 752–754, 1995.

[77] G. Zhou, Y. Chen, Z. Wang, and H. Song, "Genetic Local Search Algorithm for Optimization Design of Diffractive Optical Elements," *Applied Optics*, vol. 38, no. 20, p. 4281, 1999.

[78] K. Liu, H. Li, X. Zhang, D. Li, M. Wei, B. Li, C. Xie, and T. Zhang, "Accelerating diffractive optics design with GPU-based parallel technique," in *SPIE Optical Engineering + Applications*, pp. 778609–778609–9, International Society for Optics and Photonics, 2010.

[79] R. W. Floyd and L. Steinberg, "An adaptive algorithm for spatial greyscale," in *Sid*, vol. 17, pp. 75–77, 1976.

[80] J. Jia, Y. Wang, J. Liu, X. Li, Y. Pan, Z. Sun, B. Zhang, Q. Zhao, and W. Jiang, "Reducing the memory usage for effective computer-generated hologram calculation using compressed look-up table in full-color holographic display," *Applied Optics*, vol. 52, no. 7, pp. 1404–1412, 2013.

[81] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[82] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the Association for Computing Machinery*, vol. 42, no. 6, pp. 1115–1145, 1995.

[83] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.

[84] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM review*, vol. 57, no. 2, pp. 225–251, 2015.

[85] Y. Shechtman, Y. C. Eldar, A. Szameit, and M. Segev, "Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing," *Optics Express*, vol. 19, no. 16, pp. 14807–14822, 2011.

[86] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[87] J. Carpenter and T. D. Wilkinson, "Computation of high-stability DC balancing scheme for ferroelectric liquid crystal on silicon holograms using graphics processing units," *Optics Letters*, vol. 36, no. 8, p. 1323, 2011.

[88] D. P. Casasent and F. M. Coetzee, "Error-diffusion multilevel phase encoded CGH elements," in *San Diego'90, 8-13 July*, pp. 78–85, International Society for Optics and Photonics, 1990.

[89] T. Shimobaba, T. Kakue, and T. Ito, "Review of Fast Algorithms and Hardware Implementations on Computer Holography," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1611–1622, 2016.

[90] V. M. Bove, Q. Y. J. Smithwick, J. Barabas, and D. E. Smalley, "Is 3-D TV preparing the way for holographic TV?," in *The 8th International Symposium on Display Holography*, 2009.

[91] E. Buckley, "Computer-Generated Phase-Only Holograms for Real-Time Image Display," in *Advanced Holography - Metrology and Imaging*, pp. 277–304, InTech, 2011.

[92] J. P. Kruth, G. Levy, F. Klocke, and T. H. Childs, "Consolidation phenomena in laser and powder-bed based layered manufacturing," *CIRP Annals - Manufacturing Technology*, vol. 56, no. 2, pp. 730–759, 2007.

[93] J. Giannatsis and V. Dedoussis, "Additive fabrication technologies applied to medicine and health care: a review," *The International Journal of Advanced Manufacturing Technology*, vol. 40, no. 1-2, pp. 116–127, 2009.

[94] J. Ciurana, "Designing, prototyping and manufacturing medical devices: an overview," *International Journal of Computer Integrated Manufacturing*, vol. 27, no. 10, pp. 901–918, 2014.

[95] A. Uriondo, M. Esperon-Miguez, and S. Perinpanayagam, "The present and future of additive manufacturing in the aerospace sector: A review of important aspects," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 229, no. 11, pp. 2132–2147, 2015.

[96] G. N. Levy, R. Schindel, and J.-P. Kruth, "Rapid manufacturing and rapid tooling with layer manufacturing (lm) technologies, state of the art and future perspectives," *CIRP Annals-Manufacturing Technology*, vol. 52, no. 2, pp. 589–609, 2003.

[97] W. E. Frazier, "Metal additive manufacturing: a review," *Journal of Materials Engineering and Performance*, vol. 23, no. 6, pp. 1917–1928, 2014.

[98] S. K. Everton, M. Hirsch, P. Stravroulakis, R. K. Leach, and A. T. Clare, "Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing," *Materials & Design*, vol. 95, pp. 431–445, 2016.

[99] R. E. Samad and N. D. Vieira, "Geometrical method for determining the surface damage threshold for femtosecond laser pulses," *Laser Physics*, vol. 16, no. 2, pp. 336–339, 2006.

[100] O. Svelto, *Principles of lasers*, vol. 4. Springer, 2010.

[101] R. Martukanitz, P. Michaleris, T. A. Palmer, T. DebRoy, Z. K. Liu, R. Otis, T. W. Heo, and L. Q. Chen, "Toward an integrated computational system for describing the additive manufacturing process for metallic materials," *Additive Manufacturing*, vol. 1, pp. 52–63, 2014.

[102] G. C. Lim and W. M. Steen, "Measurement of the temporal and spatial power distribution of a high-power C02 laser beam," *Optics and Laser Technology*, vol. 14, no. June, pp. 149–153, 1982.

[103] T. a. Palmer and J. W. Elmer, "Characterisation of electron beams at different focus settings and work distances in multiple welders using the enhanced modified Faraday cup," *Science and Technology of Welding and Joining*, vol. 12, no. 2, pp. 161–174, 2007.

[104] J. W. Elmer and a. T. Teruya, "An enhanced faraday cup for rapid determination of power density distribution in electron beams," *Welding Journal*, vol. 80, no. 12, pp. 288s–295s, 2001.

[105] V. Manvatkar, A. De, and T. Debroy, "Heat transfer and material flow during laser assisted multi-layer additive manufacturing," *Journal of Applied Physics*, vol. 116, no. 12, p. 124905, 2014.

[106] C. X. Wang, T. Laoui, J. Bonse, P. J. Kruth, B. Lauwers, and L. Froyen, "Direct Selective Laser Sintering of Hard Metal Powders: Experimental Study and Simulation," *The International Journal of Advanced Manufacturing Technology*, vol. 19, no. 5, pp. 351–357, 2002.

[107] K. Mumtaz and N. Hopkinson, "Selective laser melting of Inconel 625 using pulse shaping," *Rapid Prototyping Journal*, vol. 16, no. 4, pp. 248–257, 2010.

[108] A. Vasinonta, J. L. Beuth, and M. Griffith, "Process Maps for Predicting Residual Stress and Melt Pool Size in the Laser-Based Fabrication of Thin-Walled Structures," *Journal of Manufacturing Science and Engineering*, vol. 129, no. 1, p. 101, 2007.

[109] M. J. Matthews, S. Rubenchik, G. Guss, *et al.*, "Characterization of melt-flow dynamics in selective laser melting (slm) processes," in *CLEO: Applications and Technology*, pp. ATh3A–5, Optical Society of America, 2015.

[110] S. A. Khairallah, A. T. Anderson, A. Rubenchik, and W. E. King, "Laser powder-bed fusion additive manufacturing: Physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones," *Acta Materialia*, vol. 108, pp. 36–45, 2016.

[111] M. Markl and C. Korner, "Multiscale Modeling of Powder Bed–Based Additive Manufacturing," *Annual Review of Materials Research*, vol. 46, no. 1, pp. 93–123, 2016.

[112] C. Qiu, C. Panwisawas, M. Ward, H. Basoalto, J. Brooks, and M. Attallah, "On the role of melt flow into the surface structure and porosity development during selective laser melting," *Acta Materialia*, vol. 96, pp. 72–79, 2015.

[113] V. Manvatkar, A. De, and T. DebRoy, "Spatial variation of melt pool geometry, peak temperature and solidification parameters during laser assisted additive manufacturing process," *Materials Science and Technology*, vol. 31, no. 8, pp. 924–930, 2015.

[114] M. Jeyakumar, M. Hamed, and S. Shankar, "Rheology of liquid metals and alloys," *Journal of non-Newtonian fluid mechanics*, vol. 166, no. 14-15, pp. 831–838, 2011.

[115] T. DebRoy and S. A. David, "Physical processes in fusion welding.pdf," *Reviews of Modern Physics*, vol. 67, no. 1, p. 85, 1995.

[116] I. Egry, E. Ricci, R. Novakovic, and S. Ozawa, "Surface tension of liquid metals and alloys—recent developments," *Advances in colloid and interface science*, vol. 159, no. 2, pp. 198–212, 2010.

[117] P. Sahoo, T. DebRoy, and M. J. Mcnallan, "Surface-tension of binary metal-surface-active solute systems under conditions relevant to welding metallurgy," *Metallurgical and Materials Transactions B*, vol. 19B, no. 3, pp. 483–491, 1988.

[118] P. Michaleris, "Modeling metal deposition on heat transfer analyses of additive manufacturing processes," *Finite Elements in Analysis and Design*, vol. 86, pp. 51–60, 2014.

[119] R. Nandan, T. Debroy, and H. K. D. H. Bhadeshia, "Recent Advances in Friction Stir Welding – Process, Weldment Structure and Properties," *Progress in Materials Science*, vol. 53, no. 6, pp. 980–1023, 2008.

[120] K. C. Mills, *Recommended values of thermophysical properties for selected commercial alloys*. Woodhead Publishing, 2002.

[121] B. Schoinochoritis, D. Chantzis, and K. Salonitis, "Simulation of metallic powder bed additive manufacturing processes with the finite element method: A critical review," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 231, no. 1, pp. 1–22, 2015.

[122] A. Fathi, E. Toyserkani, A. Khajepour, and M. Durali, "Prediction of melt pool depth and dilution in laser powder deposition," *Journal of Physics D: Applied Physics*, vol. 39, no. 12, pp. 2613–2623, 2006.

[123] M. Picasso, C. F. Marsden, J. W. Re, A. Frenk, and M. Rappaz, "A Simple but Realistic Model for Laser Cladding," *Metallurgical and Materials Transactions B*, vol. 25, no. April, pp. 281–291, 1994.

[124] A. Boschetto, L. Bottini, and F. Veniali, "Roughness modeling of alsi10mg parts fabricated by selective laser melting," *Journal of Materials Processing Technology*, vol. 241, pp. 154–163, 2017.

[125] E. R. Denlinger, M. Gouge, J. Irwin, and P. Michaleris, "Thermomechanical model development and in situ experimental validation of the Laser Powder-Bed Fusion process," *Additive Manufacturing*, vol. 16, pp. 73–80, 2017.

[126] T. Mukherjee, W. Zhang, and T. DebRoy, "An improved prediction of residual stresses and distortion in additive manufacturing," *Computational Materials Science*, vol. 126, pp. 360–372, 2017.

[127] N. E. Hodge, R. M. Ferencz, and J. M. Solberg, "Implementation of a thermomechanical model for the simulation of selective laser melting," *Computational Mechanics*, vol. 54, no. 1, pp. 33–51, 2014.

[128] P. P. Yuan and D. D. Gu, "Molten pool behaviour and its physical mechanism during selective laser melting of TiC/AlSi10Mg nanocomposites: simulation and experiments," *Journal of Physics D-Applied Physics*, vol. 48, no. 3, p. 35303, 2015.

[129] K. W. Thorsten Heelinga, Michael Cloots, "Melt pool simulation for the evaluation of process parameters in selective laser melting," *Additive Manufacturing*, vol. 14, pp. 116–125, 2017.

[130] S. Wen and Y. Shin, "Modeling of transport phenomena during the coaxial laser direct deposition process," *Journal of Applied Physics*, vol. 108, no. 4, p. 44908, 2010.

[131] X. He and J. Mazumder, "Transport phenomena during direct metal deposition," *Mazumder J Appl Phys vol no 5 p 053113 Mar*, vol. 053113, no. 2007, p. 53113, 2012.

[132] Y. S. Lee and W. Zhang, "Mesoscopic Simulation of Heat Transfer and Fluid Flow in Laser Powder Bed Additive Manufacturing," in *International Solid Free Form Fabrication Symposium, Austin*, pp. 1154–1165, 2015.

[133] C. Korner, A. Bauereiß, and E. Attar, "Fundamental consolidation mechanisms during selective beam melting of powders," *Modelling and Simulation in Materials Science and Engineering*, vol. 21, no. 8, p. 85011, 2013.

[134] S. A. Khairallah and A. Anderson, "Mesoscopic simulation model of selective laser melting of stainless steel powder," *Journal of Materials Processing Technology*, vol. 214, no. 11, pp. 2627–2636, 2014.

[135] J. Dawes, R. Bowerman, and R. Trepleton, "Introduction to the additive manufacturing powder metallurgy supply chain," *Johnson Matthey Technology Review*, vol. 59, no. 3, pp. 243–256, 2015.

[136] L. J., "Characterization and comparison of materials produced by Electron Beam Melting (EBM) of two different Ti–6Al–4 V powder fractions," *Journal of Materials Processing Technology*, vol. 213, no. 12, p. 2109, 2013.

[137] S. K. Brar and M. Verma, "Measurement of nanoparticles by light-scattering techniques," *TrAC Trends in Analytical Chemistry*, vol. 30, no. 1, pp. 4–17, 2011.

[138] N. Brunelli, R. Flagan, and K. Giapis, "Radial differential mobility analyzer for one nanometer particle classification," *Aerosol Science and Technology*, vol. 43, no. 1, pp. 53–59, 2009.

[139] J. A. Slotwinski, E. J. Garboczi, P. E. Stutzman, C. F. Ferraris, S. S. Watson, and M. A. Peltz, "Characterization of Metal Powders Used for Additive Manufacturing," *Journal of Research of the National Institute of Standards and Technology*, vol. 119, p. 460, 2014.

[140] A. J. Pinkerton and L. Li, "Direct additive laser manufacturing using gas- and water-atomised H13 tool steel powders," *International Journal of Advanced Manufacturing Technology*, vol. 25, no. 5-6, pp. 471–479, 2005.

[141] H. Qi, M. Azer, and A. Ritter, "Studies of Standard Heat Treatment Effects on Microstructure and Mechanical Properties of Laser Net Shape Manufactured INCONEL 718," *Metallurgical and Materials Transactions A*, vol. 40, no. 10, pp. 2410–2422, 2009.

[142] B. Liu, R. Wildman, C. Tuck, I. Ashcroft, and R. Hague, "Investigation the Effect of Particle Size Distribution on Processing Parameters Optimisation in Selective Laser Melting Process," *Sff*, no. mm, pp. 227–238, 2011.

[143] C. T. Schade, T. F. Murphy, and C. Walton, "Development of atomized powders for additive manufacturing," in *World Congress on Powder Metallurgy and Particulate Materials, PM 2014, May 18, 2014 - May 22, 2014*, pp. 215–226, 2014.

[144] W. Sames, F. Medina, W. Peter, S. Babu, and R. Dehoff, "Effect of Process Control and Powder Quality on Inconel 718 Produced Using Electron Beam Melting," in *8th International Symposium on Superalloy 718 and Derivatives*, pp. 409–423, John Wiley & Sons, Inc., 2014.

[145] H. P. Tang, M. Qian, N. Liu, X. Z. Zhang, G. Y. Yang, and J. Wang, "Effect of Powder Reuse Times on Additive Manufacturing of Ti-6Al-4V by Selective Electron Beam Melting," *JOM Journal of the Minerals Metals and Materials Society*, vol. 67, no. 3, pp. 555–563, 2015.

[146] A. Klassen, Thorsten˜Scharowsky, and Carolin˜Korner, "Evaporation model for beam based additive manufacturing using free surface lattice Boltzmann methods," *Journal of Physics D: Applied Physics*, vol. 47, no. 27, pp. 275–303, 2014.

[147] A. Ilin, R. Logvinov, A. Kulikov, A. Prihodovsky, H. Xu, V. Ploshikhin, B. Günther, and F. Bechmann, "Computer aided optimisation of the thermal management during laser beam melting process," *Physics Procedia*, vol. 56, no. C, pp. 390–399, 2014.

[148] T. Amine, "Investigation of effect of process parameters on multilayer builds by direct metal deposition," *Applied Thermal Engineering*, vol. 73, no. 1, pp. 500–511, 2015.

[149] H. Krauss, T. Zeugner, and M. F. Zaeh, "Layerwise monitoring of the selective laser melting process by thermography," *Physics Procedia*, vol. 56, pp. 64–71, 2014.

[150] J. W. Elmer, "Microstructural development during solidification of stainless steel alloys," *Metallurgical and Materials Transactions A*, vol. 20, no. 10, pp. 2117–2131, 1988.

[151] T. Mukherjee, V. Manvatkar, A. De, and T. Debroy, "Dimensionless numbers in additive manufacturing; Dimensionless numbers in additive manufacturing," *J. Appl. Phys*, vol. 121, no. 064904, p. 64904, 2017.

[152] B. E. Carroll, A. Palmer, and A. M. Beese, "ScienceDirect Anisotropic tensile behavior of Ti – 6Al – 4V components fabricated with directed energy deposition additive manufacturing," *Acta Materialia*, vol. 87, pp. 309–320, 2015.

[153] B. Cheng and K. Chou, "Melt pool evolution study in selective laser melting," in *Proceedings of the 26th Annual International Solid Freeform Fabrication Symposium*, 2015.

[154] W. Shifeng, L. Shuai, W. Qingsong, C. Yan, Z. Sheng, and S. Yusheng, "Effect of molten pool boundaries on the mechanical properties of selective laser melting parts," *Journal of Materials Processing Technology*, vol. 214, no. 11, pp. 2660–2667, 2014.

[155] Y.-L. Hung, Y.-Y. Chang, M.-J. Wang, and S.-Y. Lin, "A simple method for measuring the superhydrophobic contact angle with high accuracy," *Review of Scientific Instruments*, vol. 81, no. 6, p. 065105, 2010.

[156] P. Lott, H. Schleifenbaum, W. Meiners, K. Wissenbach, C. Hinke, and J. Bültmann, "Design of an optical system for the in situ process monitoring of selective laser melting (slm)," *Physics Procedia*, vol. 12, pp. 683–690, 2011.

[157] S. Clijsters, T. Craeghs, S. Buls, K. Kempen, and J.-P. Kruth, "In situ quality control of the selective laser melting process using a high-speed, real-time melt pool monitoring system," *The International Journal of Advanced Manufacturing Technology*, vol. 75, no. 5-8, pp. 1089–1101, 2014.

[158] S. Berumen, F. Bechmann, S. Lindner, J.-P. Kruth, and T. Craeghs, "Quality control of laser-and powder bed-based additive manufacturing (am) technologies," *Physics procedia*, vol. 5, pp. 617–622, 2010.

[159] T. Craeghs, S. Clijsters, J.-P. Kruth, F. Bechmann, and M.-C. Ebert, "Detection of process failures in layerwise laser melting with optical process monitoring," *Physics Procedia*, vol. 39, pp. 753–759, 2012.

[160] *Online quality control of selective laser melting*, 2011.

[161] T. Furumoto, M. R. Alkahari, T. Ueda, M. S. A. Aziz, and A. Hosokawa, "Monitoring of laser consolidation process of metal powder with high speed video camera," *Physics Procedia*, vol. 39, pp. 760–766, 2012.

[162] M. Pavlov, M. Doubenskaia, and I. Smurov, "Pyrometric analysis of thermal processes in slm technology," *Physics Procedia*, vol. 5, pp. 523–531, 2010.

[163] M. Doubenskaia, M. Pavlov, and Y. Chivel, "Optical System for On-Line Monitoring and Temperature Control in Selective Laser Melting Technology," in *Key Engineering Materials*, vol. 437, pp. 458–461, Trans Tech Publ, 2010.

[164] M. Islam, T. Purtonen, H. Piili, A. Salminen, and O. Nyrhilä, "Temperature profile and imaging analysis of laser additive manufacturing of stainless steel," *Physics Procedia*, vol. 41, pp. 835–842, 2013.

[165] S. Kleszczynski, J. Zur Jacobsmühlen, J. Sehrt, and G. Witt, "Error detection in laser beam melting systems by high resolution imaging," in *Proceedings of the twenty third annual international solid freeform fabrication symposium*, vol. 2012, 2012.

[166] B. Zhang, J. Ziegert, F. Farahi, and A. Davies, "In situ surface topography of laser powder bed fusion using fringe projection," *Additive Manufacturing*, vol. 12, pp. 100–107, 2016.

[167] R. J. Smith, M. Hirsch, R. Patel, W. Li, A. T. Clare, and S. D. Sharples, "Spatially resolved acoustic spectroscopy for selective laser melting," *Journal of Materials Processing Technology*, vol. 236, pp. 93–102, 2016.

[168] T. Purtonen, A. Kalliosaari, and A. Salminen, "Monitoring and adaptive control of laser processes," *Physics Procedia*, vol. 56, pp. 1218–1231, 2014.

[169] V. Giuliani, B. de Witt, M. Salluzzi, R. Hugo, and P. Gu, "Particle velocity detection in laser deposition processing," *Rapid Prototyping Journal*, vol. 14, no. 3, pp. 141–148, 2008.

[170] M. Grasso, A. Demir, B. Previtali, and B. Colosimo, "In situ monitoring of selective laser melting of zinc powder via infrared imaging of the process plume," *Robotics and Computer-Integrated Manufacturing*, vol. 49, pp. 229–239, 2018.

[171] K. Pangovski, M. Sparkes, and W. O'Neill, "A holographic method for optimisation of laser-based production processes," *Advanced Optical Technologies*, vol. 5, no. 2, pp. 177–186, 2016.

[172] K. Pangovski, O. B. Otanocha, S. Zhong, M. Sparkes, Z. Liu, W. O'Neill, and L. Li, "Investigation of plume dynamics during picosecond laser ablation of h13 steel using high-speed digital holography," *Applied Physics A*, vol. 123, no. 2, p. 114, 2017.

[173] M. Thomas, G. J. Baxter, and I. Todd, "Normalised model-based processing diagrams for additive layer manufacture of engineering alloys," *Acta Materialia*, vol. 108, pp. 26–35, 2016.

[174] V.-P. Matilainen, H. Piili, A. Salminen, and O. Nyrhilä, "Preliminary investigation of keyhole phenomena during single layer fabrication in laser additive manufacturing of stainless steel," *Physics Procedia*, vol. 78, pp. 377–387, 2015.

[175] K. Harding, *Handbook of optical dimensional metrology*. CRC Press, 2013.

[176] Q. Jia and D. Gu, "Selective laser melting additive manufacturing of Inconel 718 superalloy parts: Densification, microstructure and properties," *Journal of Alloys and Compounds*, vol. 585, pp. 713–721, 2014.

[177] H. D. Carlton, A. Haboub, G. F. Gallegos, D. Y. Parkinson, and A. A. MacDowell, "Damage Evolution and Failure Mechanisms in Additively Manufactured Stainless Steel," *Materials Science and Engineering: A*, vol. 651, pp. 406–414, 2015.

[178] M. Svensson and U. Ackelid, "Titanium Alloys Manufactured with Electron Beam Melting Mechanical and Chemical Properties," in *Medical Device Materials V: Processings from the Materials & Processes for Medical Devices Conference*, pp. 189–194, ASM International, 2009.

[179] W. E. King, H. D. Barth, V. M. Castillo, G. F. Gallegos, J. W. Gibbs, D. E. Hahn, C. Kamath, and A. M. Rubenchik, "Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing," *Journal of Materials Processing Technology*, vol. 214, no. 12, pp. 2915–2925, 2014.

[180] J. W. Elmer, J. Vaja, H. D. Carlton, and R. Pong, "The Effect of Ar and N 2 Shielding Gas on Laser Weld Porosity in Steel , Stainless Steels , and Nickel," *Welding Journal*, vol. 94, no. 10, pp. 313–325, 2015.

[181] J. A. Slotwinski, E. J. Garboczi, and K. M. Hebenstreit, "Porosity Measurements and Analysis for Metal Additive Manufacturing Process Control," *Journal of Research of the National Institute of Standards and Technology*, vol. 119, p. 494, 2014.

[182] L. Thijs, F. Verhaeghe, T. Craeghs, J. Van Humbeeck, and J.-P. Kruth, "A study of the micro structural evolution during selective laser melting of Ti-6Al-4V," *Acta Materialia*, vol. 58, no. 9, pp. 3303–3312, 2010.

[183] N. T. Aboulkhair, N. M. Everitt, I. Ashcroft, and C. Tuck, "Reducing porosity in AlSi10Mg parts processed by selective laser melting," *Additive Manufacturing*, vol. 1, pp. 77–86, 2014.

[184] X. Gong, T. Anderson, and K. Chou, "Review on Powder-Based Electron Beam Additive Manufacturing Technology," in *ASME/ISCIE 2012 International Symposium on Flexible Automation*, p. 507, American Society of Mechanical Engineers, 2012.

[185] A. Bauereiß, T. Scharowsky, and C. Körner, "Journal of Materials Processing Technology Defect generation and propagation mechanism during additive manufacturing by selective beam melting," *Journal of Materials Processing Technology*, vol. 214, no. 11, pp. 2522–2528, 2014.

[186] G. Strano, L. Hao, R. M. Everson, and K. E. Evans, "Journal of Materials Processing Technology Surface roughness analysis , modelling and prediction in selective laser melting," *Journal of Materials Processing Technology*, vol. 213, no. 4, pp. 589–597, 2013.

[187] K. Mumtaz and N. Hopkinson, "Top surface and side roughness of Inconel 625 parts processed using selective laser melting," *Rapid Prototyping Journal*, vol. 15, no. 2, pp. 96–103, 2009.

[188] A. B. Spierings, N. Herres, and G. Levy, "Influence of the particle size distribution on surface quality and mechanical properties in am steel parts," *Rapid Prototyping Journal*, vol. 17, no. 3, pp. 195–202, 2011.

[189] J.-P. Kruth, L. Froyen, J. Van Vaerenbergh, P. Mercelis, M. Rombouts, and B. Lauwers, "Selective laser melting of iron-based powder," *Journal of Materials Processing Technology*, vol. 149, no. 1, pp. 616–622, 2004.

[190] I. Todd, "Scripta Materialia Design for additive manufacturing with site-speci fi c properties in metals and alloys," *Scripta Materialia*, vol. 135, pp. 105–110, 2017.

[191] S. Kou, *Welding Metallurgy*. John Wiley & Sons, 2002.

[192] X. Zhao, X. Lin, J. Chen, L. Xue, and W. Huang, "The effect of hot isostatic pressing on crack healing, microstructure, mechanical properties of Rene88DT superalloy prepared by laser solid forming," *Materials Science and Engineering: A*, vol. 504, no. 1, pp. 129–134, 2009.

[193] P. M. J.-p. Kruth, P. Mercelis, and J.-p. Kruth, "Residual stresses in selective laser sintering and selective laser melting," *Rapid Prototyping Journal*, vol. 12, no. 5, pp. 254–265, 2010.

[194] M. Labudovic, D. Hu, and R. Kovacevic, "A three dimensional model for direct laser metal powder deposition and rapid prototyping," *Journal of Material Science*, vol. 8, no. 38, pp. 35–49, 2003.

[195] J. Smith, W. Xiong, W. Yan, S. Lin, P. Cheng, O. L. Kafka, G. J. Wagner, J. Cao, and W. K. Liu, "ME SPRING 2016 Seminar Series Linking Process , Structure , Property , and Performance for Metal-based Additive Manufacturing : Computational Approaches with Experimental Support," *Computational Mechanics*, vol. 57, no. 860, p. 2189, 2016.

[196] E. Denlinger and P. Michaleris, "Effect of stress relaxation on distortion in additive manufacturing process modeling," *Additive Manufacturing*, vol. 12, pp. 51–59, 2016.

[197] A. De and T. DebRoy, "A perspective on residual stresses in welding," *Science and Technology of Welding and Joining*, vol. 16, no. 3, pp. 204–208, 2011.

[198] T. Mukherjee, V. Manvatkar, A. De, and T. DebRoy, "Mitigation of thermal distortion during additive manufacturing," *Scripta Materialia*, vol. 127, pp. 79–83, 2017.

[199] M. Alimardani, E. Toyserkani, and J. Huissoon, "A 3D dynamic numerical approach for temperature and thermal stress distributions in multilayer laser solid freeform fabrication process," *Optics and Lasers in Engineering*, vol. 45, no. 12, pp. 1115–1130, 2007.

[200] Y. Chew, J. H. L. Pang, G. Bi, and B. Song, "Thermo-mechanical model for simulating laser cladding induced residual stresses with single and multiple clad beads," *Journal of Materials Processing Technology*, vol. 224, pp. 89–101, 2015.

[201] E. R. Denlinger, J. Irwin, and P. Michaleris, "Thermomechanical Modeling of Additive Manufacturing Large Parts," *Journal of Manufacturing Science and Engineering*, vol. 136, no. 6, p. 061007, 2014.

[202] T. Mukherjee, J. S. Zuback, A. De, and T. DebRoy, "Printability of alloys for additive manufacturing," *Scientific Reports*, vol. 6, 2016.

[203] C. Brice, R. Shenoy, M. Kral, and K. Buchannan, "Precipitation behavior of aluminum alloy 2139 fabricated using additive manufacturing," *Materials Science and Engineering A*, vol. 648, pp. 9–14, 2015.

[204] J. Perepezko and G. Wilde, "Melt undercooling and nucleation kinetics," *Current Opinion in Solid State and Materials Science*, vol. 20, no. 1, pp. 3–12, 2016.

[205] X. He, T. Debroy, and P. W. Fuerschbach, "Alloying element vaporization during laser spot welding of stainless steel," *J. Phys. D: Appl. Phys. J. Phys. D: Appl. Phys*, vol. 36, no. 36, pp. 3079–3088, 2003.

[206] H. Zhao and T. Debroy, "Weld metal composition change during conduction mode laser welding of aluminum alloy 5182," *Metallurgical and Materials Transactions B*, vol. 32, no. 1, pp. 163–172, 2001.

[207] Cuda, "www.developer.nvidia.com/cuda-zone," accessed on: 2020-12-07.

[208] P. Steinbach and M. Werner, "gearshifft–the fft benchmark suite for heterogeneous platforms," in *International Supercomputing Conference*, pp. 199–216, Springer, 2017.

[209] T. Shimobaba, N. Masuda, and T. Ito, "Simple and fast calculation algorithm for computer-generated hologram with wavefront recording plane," *Optics Letters*, vol. 34, no. 20, p. 3133, 2009.

[210] T. Shimobaba, J. Weng, T. Sakurai, N. Okada, T. Nishitsuji, N. Takada, A. Shiraki, N. Masuda, and T. Ito, "Computational wave optics library for C++: CWO++ library," *Computer Physics Communications*, vol. 183, no. 5, pp. 1124–1138, 2012.

[211] Y. Wang, X. Sang, Z. Chen, H. Li, and L. Zhao, "Real-time photorealistic computer-generated holograms based on backward ray tracing and wavefront recording planes," *Optics Communications*, vol. 429, no. July, pp. 12–17, 2018.

[212] N. Takada, T. Shimobaba, H. Nakayama, A. Shiraki, N. Okada, M. Oikawa, N. Masuda, and T. Ito, "Fast high-resolution computer-generated hologram computation using multiple graphics processing unit cluster system," *Applied Optics*, vol. 51, no. 30, p. 7303, 2012.

[213] T. Shimobaba and T. Ito, "Fast generation of computer-generated holograms using wavelet shrinkage," *Optics Express*, vol. 25, no. 1, p. 77, 2017.

[214] T. Shimobaba, K. Matsushima, T. Takahashi, Y. Nagahama, S. Hasegawa, M. Sano, R. Hirayama, T. Kakue, and T. Ito, "Fast, large-scale hologram calculation in wavelet domain," *Optics Communications*, vol. 412, no. August 2017, pp. 80–84, 2018.

[215] T. Shimobaba, S. Yamada, T. Kakue, and T. Ito, "Fast hologram calculation using wavelet transform," in *Proc. SPIE 10964, Tenth International Conference on Information Optics and Photonics*, vol. 10964, p. 116, SPIE, 2018.

[216] K. Matsushima and S. Nakahara, "Extremely high-definition full-parallax computer-generated hologram created by the polygon-based method," *Applied Optics*, vol. 48, no. 34, pp. H54–H63, 2009.

[217] S.-C. Kim and E.-S. Kim, "Effective generation of digital holograms of three-dimensional objects using a novel look-up table method," *Applied Optics*, vol. 47, no. 19, p. D55, 2008.

[218] S.-C. Kim and E.-S. Kim, "Fast computation of hologram patterns of a 3D object using run-length encoding and novel look-up table methods," *Applied Optics*, vol. 48, no. 6, p. 1030, 2009.

[219] J. Jia, Y. Wang, J. Liu, X. Li, Y. Pan, Z. Sun, B. Zhang, Q. Zhao, and W. Jiang, "Reducing the memory usage for effectivecomputer-generated hologram calculation using compressed look-up table in full-color holographic display," *Applied Optics*, vol. 52, no. 7, p. 1404, 2013.

[220] G. Makey, M. S. El-Daher, and K. Al-Shufi, "Accelerating the calculations of binary detour phase method by integrating both CUDA and Matlab programming for GPU's parallel computations," *Optik*, vol. 124, no. 22, pp. 5486–5488, 2013.

[221] P. Memmolo, L. Miccio, F. Merola, A. Paciello, V. Embrione, S. Fusco, P. Ferraro, and P. Antonio Netti, "Investigation on specific solutions of Gerchberg-Saxton algorithm," *Optics and Lasers in Engineering*, vol. 52, no. 1, pp. 206–211, 2014.

[222] Z. Wang, G. Lv, Q. Feng, A. Wang, and H. Ming, "Highly efficient calculation method for computer-generated holographic stereogram using a lookup table," *Applied Optics*, vol. 58, no. 5, pp. A41–A47, 2019.

[223] J. Carpenter and T. D. Wilkinson, "Graphics processing unit–accelerated holography by simulated annealing," *Optical Engineering*, vol. 49, no. 9, p. 095801, 2010.

[224] MaterialSkin Library, "www.github.com/IgnaceMaes/MaterialSkin," accessed on: 2020-12-07.

[225] LiveCharts Library, "www.lvcharts.net," accessed on: 2020-12-07.

[226] SharpDX Library, "www.sharpdx.org," accessed on: 2020-12-07.

[227] Helix Toolkit Library, "www.github.com/helix-toolkit/helix-toolkit," accessed on: 2020-12-07.

[228] ControlzEx Library, "www.github.com/ControlzEx/ControlzEx," accessed on: 2020-12-07.

[229] NHotkey Library, "www.github.com/thomaslevesque/NHotkey," accessed on: 2020-12-07.

[230] Math.NET Library, "www.mathdotnet.com," accessed on: 2020-12-07.

[231] Newtonsoft.Json Library, "www.newtonsoft.com/json," accessed on: 2020-12-07.

[232] NUnit Library, "www.nunit.org," accessed on: 2020-12-07.

[233] AForge.NET Library, "www.aforgenet.com/framework," accessed on: 2020-12-07.

[234] ManagedCuda Library, "www.kunzmi.github.io/managedCuda," accessed on: 2020-12-07.

[235] Accord.NET Library, "accord-framework.net/," accessed on: 2020-12-07.

[236] csmatio Library, "sourceforge.net/projects/csmatio/," accessed on: 2020-12-07.

[237] PdfiumViewer Library, "github.com/pvginkel/PdfiumViewer," accessed on: 2020-12-07.

[238] CefSharp Library, "github.com/cefsharp/CefSharpr," accessed on: 2020-12-07.

[239] ClosedXML Library, "github.com/closedxml/closedxml," accessed on: 2020-12-07.

[240] MoreLinq Library, “github.com/morelinq/MoreLINQ,” accessed on: 2020-12-07.

[241] GongSolutions.WPF.DragDrop Library, “github.com/punker76/gong-wpf-dragdrop,” accessed on: 2020-12-07.

[242] MaterialSkin Library License, “www.github.com/IgnaceMaes/MaterialSkin/blob/master/LICENSE,” accessed on: 2020-12-07.

[243] LiveCharts Library License, “www.github.com/beto-rodriguez/Live-Charts/blob/master/LICENSE.TXT,” accessed on: 2020-12-07.

[244] SharpDX Library License, “www.github.com/sharpdx/SharpDX/blob/master/License.txt,” accessed on: 2020-12-07.

[245] Helix Toolkit Library License, “www.github.com/helix-toolkit/helix-toolkit/blob/develop/LICENSE,” accessed on: 2020-12-07.

[246] ControlzEx Library License, “www.github.com/ControlzEx/ControlzEx/blob/develop/LICENSE,” accessed on: 2020-12-07.

[247] Math.NET Library License, “www.github.com/mathnet/mathnet-numerics/blob/master/LICENSE.md,” accessed on: 2020-12-07.

[248] Newtonsoft.Json Library License, “www.github.com/JamesNK/Newtonsoft.Json/blob/master/LICENSE.md,” accessed on: 2020-12-07.

[249] NUnit Library License, “www.nunit.org/nuget/nunit3-license.txt,” accessed on: 2020-12-07.

[250] MIT License, “www.opensource.org/licenses/MIT,” accessed on: 2020-12-07.

[251] ClosedXML Library License, “github.com/ClosedXML/ClosedXML/blob/develop/LICENSE,” accessed on: 2020-12-07.

[252] AForge.NET Library License, "www.aforgenet.com/framework/license.html," accessed on: 2020-12-07.

[253] ManagedCuda Library License, "www.github.com/kunzmi/managedCuda/blob/master/LICENSE.txt," accessed on: 2020-12-07.

[254] Accord.NET Library License, "github.com/accord-net/framework/blob/development/LICENSE," accessed on: 2020-12-07.

[255] Lesser General Public License, "www.gnu.org/licenses/lgpl-3.0.en.html," accessed on: 2020-12-07.

[256] NHotkey Library License, "www.github.com/thomaslevesque/NHotkey/blob/master/LICENSE.md," accessed on: 2020-12-07.

[257] PdfiumViewer Library License, "github.com/pvginkel/PdfiumViewer/blob/master/LICENSE," accessed on: 2020-12-07.

[258] Apache License 2.0, "www.apache.org/licenses/LICENSE-2.0," accessed on: 2020-12-07.

[259] CefSharp Library License, "github.com/cefsharp/CefSharp/blob/master/LICENSE," accessed on: 2020-12-07.

[260] The 3-Clause BSD License, "opensource.org/licenses/BSD-3-Clause," accessed on: 2020-12-07.

[261] GongSolutions.WPF.DragDrop Library License, "github.com/punker76/gong-wpf-dragdrop/blob/dev/LICENSE," accessed on: 2020-12-07.

[262] Cuda EULA, "docs.nvidia.com/cuda/eula/index.html," accessed on: 2020-12-07.

[263] Xamarin.Forms, "www.xamarin.com/forms," accessed on: 2020-12-07.

[264] D. Blinder, A. Ahar, S. Bettens, T. Birnbaum, A. Symeonidou, H. Ottevaere, C. Schretter, and P. Schelkens, "Signal processing challenges for digital holographic

video display systems," *Signal Processing: Image Communication*, vol. 70, pp. 114–130, 2019.

[265] I. of Electrical and E. Engineers, "IEEE Standard for Binary Floating-Point Arithmetic," standard, Institute of Electrical and Electronics Engineers, 2008.

[266] P. J. Christopher, J. D. Lake, D. Dong, H. J. Joyce, and T. D. Wilkinson, "Improving holographic search algorithms using sorted pixel selection," *Journal of the Optical Society of America A*, vol. 36, pp. 1456–1462, Sep 2019.

[267] P. J. Christopher, R. Mouthaan, V. Bheemireddy, and T. D. Wilkinson, "Improving performance of single-pass real-time holographic projection," *Optics Communications*, vol. 457, p. 124666, 2020.

[268] P. J. Christopher, Y. Wang, and T. D. Wilkinson, "Novel predictive search algorithm for phase holography," *Journal of the Optical Society of America A*, N.D. Journal of the Optical Society of America A, In Production.

[269] P. J. Christopher, R. Mouthaan, G. S. Gordon, and T. D. Wilkinson, "Holographic predictive search: Extending the scope," *Optics Communications*, vol. 467, p. 125701, 2020.

[270] P. J. Christopher, R. Mouthaan, A. M. Soliman, and T. D. Wilkinson, "Sympathetic quantisation - a new approach to hologram quantisation," *Optics Communications*, vol. 473, p. 125883, 2020.

[271] P. J. Christopher, R. Mouthaan, M. E. Guendy, and T. D. Wilkinson, "Linear-time algorithm for phase-sensitive holography," *Optical Engineering*, vol. 59, no. 8, pp. 1 – 10, 2020.

[272] D. E. Knuth, *The art of computer programming: sorting and searching*, vol. 3. Pearson Education, 1997.

[273] W. J. Dallas, "Magnitude-coupled phase quantization," *Applied Optics*, vol. 13, pp. 2274–2279, Oct. 1974.

[274] K. D. Rines and N. C. Gallagher, "Reducing quantization error in hsueh-sawchuk holograms," *Applied Optics*, vol. 20, pp. 2008–2010, June 1981.

[275] C. K. Hsueh and A. A. Sawchuk, "Computer-generated double-phase holograms," *Applied Optics*, vol. 17, pp. 3874–3883, Dec. 1978.

[276] H. Akahori, "Spectrum leveling by an iterative algorithm with a dummy area for synthesizing the kinoform," *Applied Optics*, vol. 25, pp. 802–811, Mar. 1986.

[277] J. R. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," *Optical Engineering*, vol. 19, no. 3, pp. 297 – 305 – 9, 1980.

[278] S. Yang and T. Shimomura, "Error dependence of quantized kinoform reconstruction on the position of the desired image," *Applied Optics*, vol. 39, pp. 2896–2903, June 2000.

[279] G. . Kang, J. . Xie, X. . Mo, D. . Wang, and H. Zhang, "Design of binary optics using improved two-step simulated annealing algorithm," *Guangzi Xuebao/Acta Photonica Sinica*, vol. 37, no. 7, pp. 1416–1419, 2008. Cited By :5.

[280] E. Zhang, S. Noehte, C. H. Dietrich, and R. Männer, "Gradual and random binarization of gray-scale holograms," *Applied Optics*, vol. 34, no. 26, pp. 5987–5995, 1995. Cited By :25.

[281] J. Amako, H. Miura, and T. Sonehara, "Speckle-noise reduction on kinoform reconstruction using a phase-only spatial light modulator," *Applied Optics*, vol. 34, no. 17, pp. 3165–3171, 1995.

[282] Y. Takaki and M. Yokouchi, "Speckle-free and grayscale hologram reconstruction using time-multiplexing technique," *Optics Express*, vol. 19, no. 8, pp. 7567–7579, 2011.

[283] S.-J. Liu, D. Wang, and Q.-H. Wang, "Speckle noise suppression method in holographic display using time multiplexing technique," *Optics Communications*, vol. 436, pp. 253 – 257, 2019.

[284] T. Shimobaba, A. Shiraki, N. Masuda, and T. Ito, "An electroholographic colour reconstruction by time division switching of reference lights," *Journal of Optics A: Pure and Applied Optics*, vol. 9, pp. 757–760, jul 2007.

[285] T. Shimobaba and T. Ito, "A color holographic reconstruction system by time division multiplexing with reference lights of laser," *Optical Review*, vol. 10, pp. 339–341, Sep 2003.

[286] M. Oikawa, T. Shimobaba, T. Yoda, H. Nakayama, A. Shiraki, N. Masuda, and T. Ito, "Time-division color electroholography using one-chip rgb led and synchronizing controller," *Opt. Express*, vol. 19, pp. 12008–12013, Jun 2011.

[287] R. Dosselmann and X. D. Yang, "A comprehensive assessment of the structural similarity index," *Signal, Image and Video Processing*, vol. 5, pp. 81–91, Mar 2011.

[288] L. Calvet, P. Gurdjos, C. Griwodz, and S. Gasparini, "Detection and Accurate Localization of Circular Fiducials under Highly Challenging Conditions," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, United States), pp. 562 – 570, June 2016.

[289] Slic3r, "www.slic3r.org/," accessed on: 2020-12-07.

[290] J. H. Moore, C. C. Davis, M. A. Coplan, and S. C. Greer, *Building scientific apparatus*. Cambridge University Press, 2009.

[291] Y. Wang, D. Dong, P. J. Christopher, A. Kadis, R. Mouthaan, F. Yang, and T. D. Wilkinson, "Hardware implementations of computer-generated holography: a review," *Optical Engineering*, vol. 59, no. 10, pp. 1 – 30, 2020.

[292] P. M. Hirsch, J. A. Jordan Jr, and L. B. Lesem, "Method of making an object dependent diffuser," 1971.

[293] F. Wyrowski and O. Bryngdahl, "Iterative Fourier-transform algorithm applied to computer holography," *Journal of the Optical Society of America A*, vol. 5, no. 7, p. 1058, 1988.

[294] G. Sinclair, J. Leach, P. Jordan, G. Gibson, E. Yao, Z. Laczik, M. Padgett, and J. Courtial, "Interactive application in holographic optical tweezers of a multi-plane Gerchberg-Saxton algorithm for three-dimensional light shaping.," *Optics Express*, vol. 12, no. 8, pp. 1665–1670, 2004.

[295] M. Makowski, M. Sypek, A. Kolodziejczyk, and G. Mikuła, "Three-plane phase-only computer hologram generated with iterative Fresnel algorithm," *Optical Engineering*, vol. 44, no. 12, pp. 125805–125805–7, 2005.

[296] C. Chang, J. Xia, and Y. Jiang, "Holographic image projection on tilted planes by phase-only computer generated hologram using fractional fourier Transformation," *IEEE/OSA Journal of Display Technology*, vol. 10, no. 2, pp. 107–113, 2014.

[297] R. D. Leonardo, F. Ianni, and G. Ruocco, "Computer generation of optimal holograms for optical trap arrays," *Opt. Express*, vol. 15, pp. 1913–1922, Feb 2007.

[298] W.-H. Lee, "Binary computer-generated holograms," *Applied Optics*, vol. 18, pp. 3661–3669, Nov 1979.

[299] H. Kim, C.-Y. Hwang, K.-S. Kim, J. Roh, W. Moon, S. Kim, B.-R. Lee, S. Oh, and J. Hahn, "Anamorphic optical transformation of an amplitude spatial light modulator to a complex spatial light modulator with square pixels," *Applied Optics*, vol. 53, pp. G139–G146, Sep 2014.

[300] J. A. Davis, D. M. Cottrell, J. Campos, M. J. Yzuel, and I. Moreno, "Encoding amplitude information onto phase-only filters," *Applied Optics*, vol. 38, pp. 5004–5013, Aug 1999.

[301] F. Wyrowski, "Diffraction efficiency of analog and quantized digital amplitude holograms: analysis and manipulation," *Journal of the Optical Society of America A*, vol. 7, pp. 383–393, Mar 1990.

[302] T. Shimobaba and T. Ito, "Random phase-free computer-generated hologram," *Optics Express*, vol. 23, no. 7, pp. 9549–9554, 2015.

[303] D. McGraw, "Interface design for a fast ferroelectric display," Masters Thesis, University of Cambridge, 2018.

[304] "Holoblade: An open-hardware spatial light modulator driver platform for holographic displays," *Applied Optics*.

[305] Displaytech, Inc., "Customer Product Specification, Holographic Display Panel, Model: HDP-1280-2," tech. rep., Displaytech, Inc., 2007.

[306] International Organization for Standardization, "Lasers and laser-related equipment - Determination of laser-induced damage threshold of optical surfaces," standard, International Organization for Standardization, Geneva, CH, Mar. 2000.

[307] International Organization for Standardization, "Lasers and laser-related equipment - Determination of laser-induced damage threshold of optical surfaces," standard, International Organization for Standardization, Geneva, CH, Mar. 2011.

[308] R. Wood, "Laser induced damage thresholds and laser safety levels. do the units of measurement matter?," *Optics & Laser Technology*, vol. 29, no. 8, pp. 517 – 522, 1998.

[309] D. Ristau, *Laser-induced damage in optical materials*. CRC Press, 1st. ed., 2014.

[310] N. Bloembergen, "Role of cracks, pores, and absorbing inclusions on laser induced damage threshold at surfaces of transparent dielectrics," *Applied Optics*, vol. 12, pp. 661–664, Apr 1973.

[311] C. W. Carr, H. B. Radousky, and S. G. Demos, "Wavelength dependence of laser-induced damage: Determining the damage initiation mechanisms," *Phys. Rev. Lett.*, vol. 91, p. 127402, Sep 2003.

[312] S. L. U. Limited, *SP-200C-0001 200W CW/Modulated High Power Fiber Laser Product Manual, edition G*, 2006.

[313] A. Gjurchinovski, "Polarization measurements - a numerical approach," *arXiv preprint physics/0401105*, 2004.

[314] J. Peatross and M. Ware, *Physics of light and optics*. Brigham Young University, Department of Physics, 2011.

Appendix A

Variance and Error in One-Step Phase-Retrieval

Appendix A

Variance and Error in One-Step Phase-Retrieval[1]

A.1 Introduction

The rise of mixed reality systems has seen a resurgence of interest in real-time CGH and in the algorithms required. In this appendix the statistical properties of the OSPR algorithm are discussed in greater detail. Firstly, relationships for the variance are derived based on earlier work by Cable and Buckley. These are then extended to show the effect on mean squared error (MSE) and structural similarity index (SSIM), Section 2.4.1. The implications on generation performance and frame-rate is then discussed and recommendations made before conclusions are drawn.

A.2 OSPR Variance

In their proceedings paper [43] Cable and Buckley demonstrate that the variance in noise due to N subframes is proportional to the reciprocal of N. To demonstrate this it is observed

[1]The contents of this section are currently included in the following preprint:

Peter J. Christopher and Timothy D. Wilkinson, “Variance and Error in One-Step Phase-Retrieval", *arXiv*, 2019, arXiv: 1911.00045

that the perceived intensity of a subframe is given as the square of the replay field amplitude $R\overline{R}$ where R is the amplitudes of the replay field given by the Fourier transform of hologram H. For N subframes the perceived intensity is equal to

$$I_{u,v} = R_{u,v}\overline{R_{u,v}} = \frac{1}{N}\sum_{n=1}^{N} \mathscr{F}(H_n)_{u,v}\overline{\mathscr{F}(H_n)_{u,v}} \tag{A.1}$$

$$\forall \quad u \in \mathbb{Z} \cup (0, N_x] \wedge v \in \mathbb{Z} \cup (0, N_y] \tag{A.2}$$

It is assumed that the value of $R\overline{R}$ is equal to the target image intensities $T\overline{T}$ plus a noise term ε' with mean $\mu_{\varepsilon'}$ and variance $\sigma^2_{\varepsilon'}$ where ε' is a complex random variable of circularly symmetric distribution. It is also assumed that Parseval's theorem holds and that the total energy in the diffraction and replay fields is the same.

For a single sub-frame system this leads to

$$R_{u,v}\overline{R_{u,v}} = T_{u,v}\overline{T_{u,v}} + \varepsilon'_{u,v}. \tag{A.3}$$

For systems utilising N sub-frames this becomes.

$$R_{u,v}\overline{R_{u,v}} = T_{u,v}\overline{T_{u,v}} + \frac{1}{N}\sum_{n=1}^{N} \varepsilon'_{n,u,v}. \tag{A.4}$$

Provided N is sufficiently large and ε is independent and identically distributed (i.i.d.) for all n, u, v, the central limit theorem (CLT) can be applied. Leading to

$$R_{u,v}\overline{R_{u,v}} = T_{u,v}\overline{T_{u,v}} + \varepsilon_{u,v}. \tag{A.5}$$

where random variable ε is a bivariate Gaussian random variable with mean $\mu_\varepsilon = \mu_{\varepsilon'}$ and variance $\sigma^2_\varepsilon = \sigma^2_{\varepsilon'}/N$. This means that, provided the error terms are independent, the error variance will be equal to the reciprocal of the number of frames.

It is worth noting that the number of sub-frames is anticipated to be low in real world systems and that there will be noticeable non-linearity in the variance due to the CLT.

A.3 OSPR Mean Squared Error

It is tempting to think that, provided that energy conservation is upheld, that the mean squared error $\mu_\varepsilon = 0$ will be equal to the variance. Unfortunately, due to the no this does not follow for reasons discussed individually here.

As the system is non-linear, the MSE - Section 2.4.1 will actually consist of a variance term summed with a constant bias term. Two common cases cause the bias

A.3.1 Conjugate Symmetry

The first reason for bias is due to the conjugate image symmetry requirements in binary devices which mandates that a replay field must be equal to itself when subject to a 180° rotation around the centre of the field. i.e. for a pixel $R_{u,v}$ on a replay field of size N_x, N_y the replay field values will follow the relationship

$$R_{u,v} \equiv R_{N_x-u,N_y-v} \quad \forall \quad u \in \mathbb{Z} \cup (0,N_x] \wedge v \in \mathbb{Z} \cup (0,N_y]\,. \tag{A.6}$$

This symmetry relationship fails the i.i.d. requirements. Instead of (A.5), the following applies

$$\begin{aligned} R_{u,v}\overline{R_{u,v}} &= R_{N_x-u,N_y-v}\overline{R_{N_x-u,N_y-v}} \\ &= \frac{T_{u,v}\overline{T_{u,v}} + \varepsilon_{u,v} + T_{N_x-u,N_y-v}\overline{T_{N_x-u,N_y-v}} + \varepsilon_{N_x-u,N_y-v}}{2} \end{aligned} \tag{A.7}$$

which leads to an equation for the bias $\mathrm{Bias}_{cs}(T,R)$ due to conjugate symmetry of

$$\begin{aligned} &\mathrm{Bias}_{cs}(T,R)^2 = \\ &\frac{1}{N_xN_y}\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left(\frac{\left|R_{u,v}\overline{R_{u,v}} - R_{N_x-u,N_y-v}\overline{R_{N_x-u,N_y-v}}\right|}{2}\right) \end{aligned} \tag{A.8}$$

For rotationally symmetric target images, this is expected to be zero.

While this symmetry requirement is well understood, more involved symmetry requirements exist for almost every binary device. For example, a binary phase device not modulating to an interval of π will introduce complicated periodic symmetry requirements.

A.3.2 Intensity Distribution

The second source of bias is due to the nature of the noise distribution. While the noise variance converges to zero as $N \to \infty$, the intensity is determined by the square of the replay amplitudes. This translates to the mean of the image intensities not being linearly related to the mean of the image amplitudes.

If ε is taken as being circularly symmetric and normally distributed then (A.5) has a Rician probability density function (PDF)

$$p_{R\overline{R}}(R_{u,v}\overline{R_{u,v}}) = \frac{R_{u,v}\overline{R_{u,v}}}{\sigma_{\varepsilon'}^2} \exp\left(-\frac{(R_{u,v}\overline{R_{u,v}})^2 + (T_{u,v}\overline{T_{u,v}})^2)}{2\sigma_{\varepsilon'}^2}\right) J_0\left(\frac{R_{u,v}\overline{R_{u,v}}T_{u,v}\overline{T_{u,v}}}{\sigma_{\varepsilon'}^2}\right), \tag{A.9}$$

where J_0 is the first Bessel function. In the limit as ${}^{T_{u,v}\overline{T_{u,v}}}/_{\sigma_{\varepsilon'}} \to \infty$, this tends to the Gaussian distribution while as ${}^{T_{u,v}\overline{T_{u,v}}}/_{\sigma_{\varepsilon'}} \to 0$ this tends to the Rayleigh distribution.

If the target magnitudes are taken as having an amplitude distribution of $p_{T\overline{T}}(T_{u,v}\overline{T_{u,v}})$ then can write an integral for the bias $\mathrm{Bias}_{id}(T,R)$ due to the intensity distribution of

$$\mathrm{Bias}_{id}(T,R)^2 = \int_0^\infty p_{T\overline{T}}(T_{u,v}\overline{T_{u,v}}) \int_0^\infty p_{R\overline{R}}(R_{u,v}\overline{R_{u,v}}) \left(R_{u,v}\overline{R_{u,v}} - T_{u,v}\overline{T_{u,v}}\right)^2 d(R_{u,v}\overline{R_{u,v}}) d(T_{u,v}\overline{T_{u,v}}) \tag{A.10}$$

Analytical solutions are possible but complex. Numerical solutions are practical however. The key observation is that while variance σ_ε of perceived intensity over N subframes decreases as the reciprocal of N the mean $\mu_{\mathrm{id},\varepsilon}^2$ of perceived intensity over N subframes remains a function of $\sigma_{\varepsilon'}$.

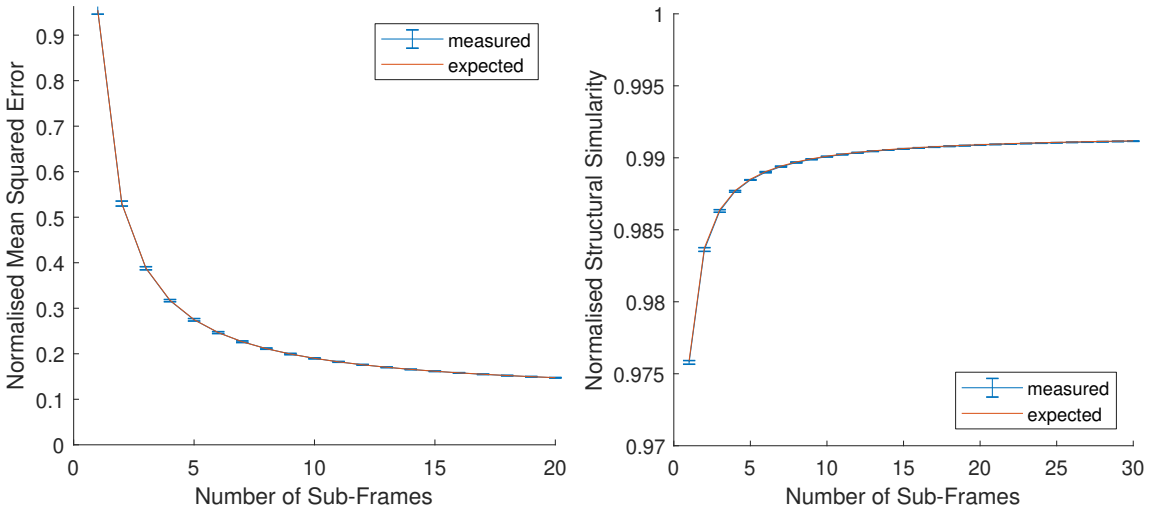


Fig. A.1 Time averaged mean squared error (left) and structural similarity (right) for OSPR.

In practice regression may be used. For example Figure A.1 (left) shows the convergence of the MSE against number of subframes for a modified *Mandrill* test image which provides a close fit to the expected $E = A + {}^{B}/_{N}$ curve. In Figure A.1 the bias $\text{Bias}_{id}(T,R)$ is equal to 0.068 with initial variance of $\sigma_{\varepsilon'}^2$ equal to 0.884. Values are taken as being the mean of 100 independent runs with error bars showing two standard deviations. The 512×512 pixel *Mandrill* test image has an artificially induced symmetry and is modelled for a binary phase SLM.

$$Error(T,R) = \text{Bias}_{id}(T,R)^2 + \sigma_{\varepsilon}^2 \tag{A.11}$$

As both $\text{Bias}_{id}(T,R)^2$ and σ_{ε}^2 are determined purely by the intensity distribution for images i.i.d. in n, u and v this convergence graph will be equal for similarly distributed images. An interesting observation of this is that uniformly distributed image magnitudes will have a higher bias than Gaussian distributions

A.3.3 Performance

Table A.1 shows examples of the bias and variance for selected amplitude distributions. As expected from our earlier observations, amplitude distributions with more terms near 0 had a larger bias term than those with amplitude terms with fewer terms near from 0. In the most extreme case, constant amplitude, the bias term was negligible. Of especial interest is that

Table A.1 Mean squared error parameters from (A.11) for different amplitude distributions and uniform $[-\pi, \pi)$ phase distribution.

Amplitude	Measured MSE		Simulated MSE	
Distribution	$\text{Bias}_{id}(T,R)$	$\sigma^2_{\varepsilon'}$	$\text{Bias}_{id}(T,R)$	$\sigma^2_{\varepsilon'}$
Uniform	0.452	0.797	0.455	0.794
Constant	0.001	0.799	0.000	0.797
Mandrill	0.068	0.884	0.067	0.883
Peppers	0.069	0.883	0.066	0.884

natural images such as *Mandrill* and *Peppers* which have a more central distribution had a smaller bias than uniformly distributed images.

Table A.1 also shows *measured* and *simulated* values. Measured values were determined by taking the mean and deviations of running OSPR for a given number of subframes. Simulated values were taken by numerically integrating (A.10) for the given distribution of amplitudes. The simulated values were accurate to within 1% of the measured values.

A.4 OSPR Structural Similarity Index

In Section A.2 relationship between variance and number of subframes for the OSPR algorithm was discussed. In Section A.3 a similar relationship for the bias term given in (A.10) was developed. From this the expected MSE for a given distribution of amplitudes can be determined.

As discussed in Section 2.4.1 for human eye applications, SSIM is more commonly used than MSE as it has been shown to more closely correspond to ocular visual quality[48].

Figure A.2 shows the components of this relationship for a uniformly distributed amplitude image for 1 and 10 subframes. As expected μ_T^2 and σ_T^2 do not change with time. The mean value for μ_R^2 stays as 1 but the summing effect of more subframes means that the distribution becomes narrower. The mean value for σ_R^2 follows the expected $1/N$ relationship and can be assumed identical to $\sigma^2_{\varepsilon'}$ for large windows.

It can also be seen from Figure A.2 that s_2 is more prominent than s_1 as a factor. If $s_1 =$ is set to 1 then

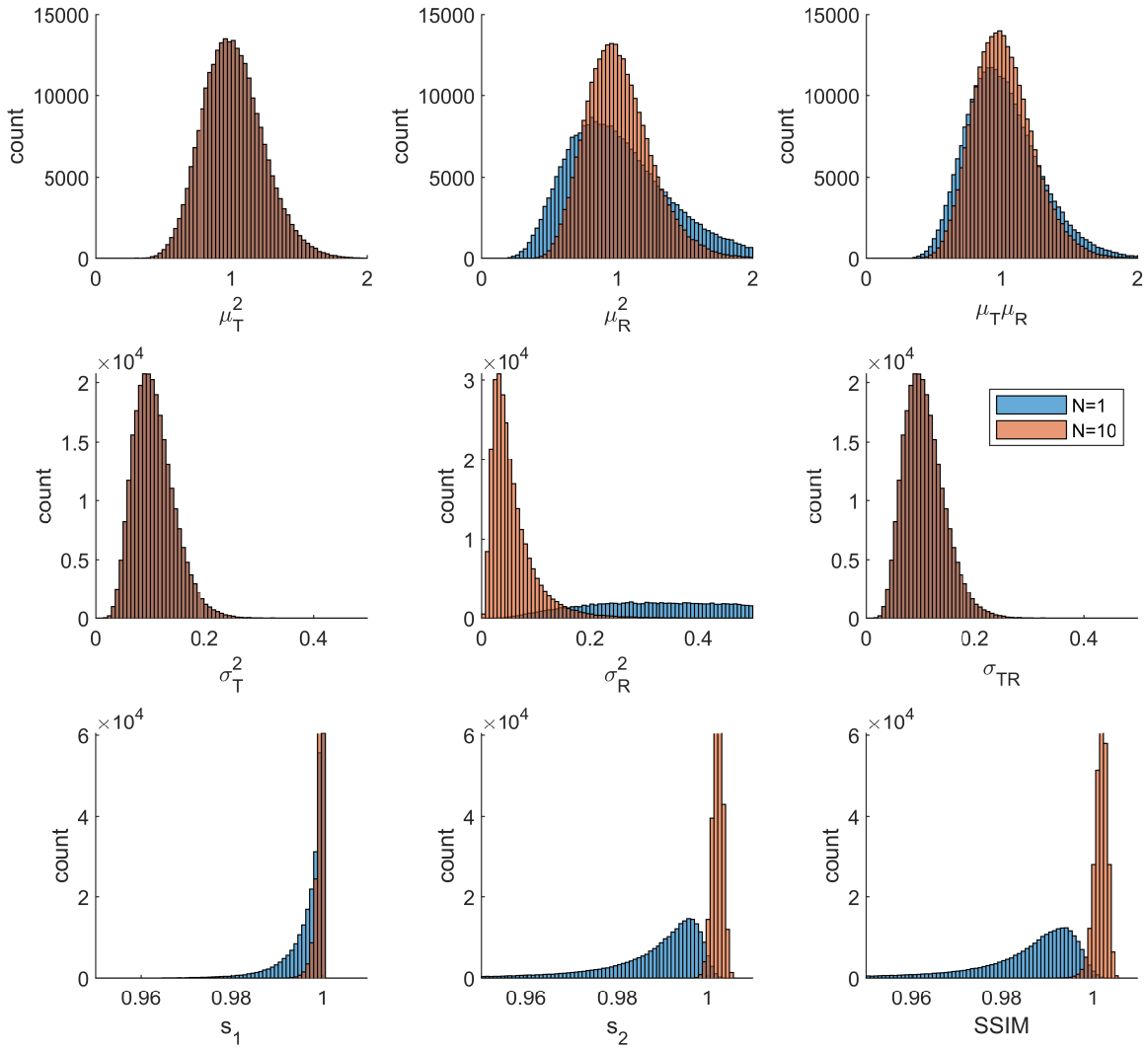


Fig. A.2 Evolution of components of the SSIM equation over time. Image used is a 512×512 uniformly distributed amplitude.

$$SSIM(T,R) \approx \frac{c_2}{\left(2\sigma_T^2 + \sigma_{\varepsilon'}^2/N + c_2\right)} \tag{A.12}$$

Figure A.3 shows the same components for the *Mandrill* test image. While the distributions have changed, the behaviour and properties can be seen to be similar.

By integrating (A.12) numerically the result shown in Figure A.1 (right) is obtained which provides a close fit to the measured behaviour.

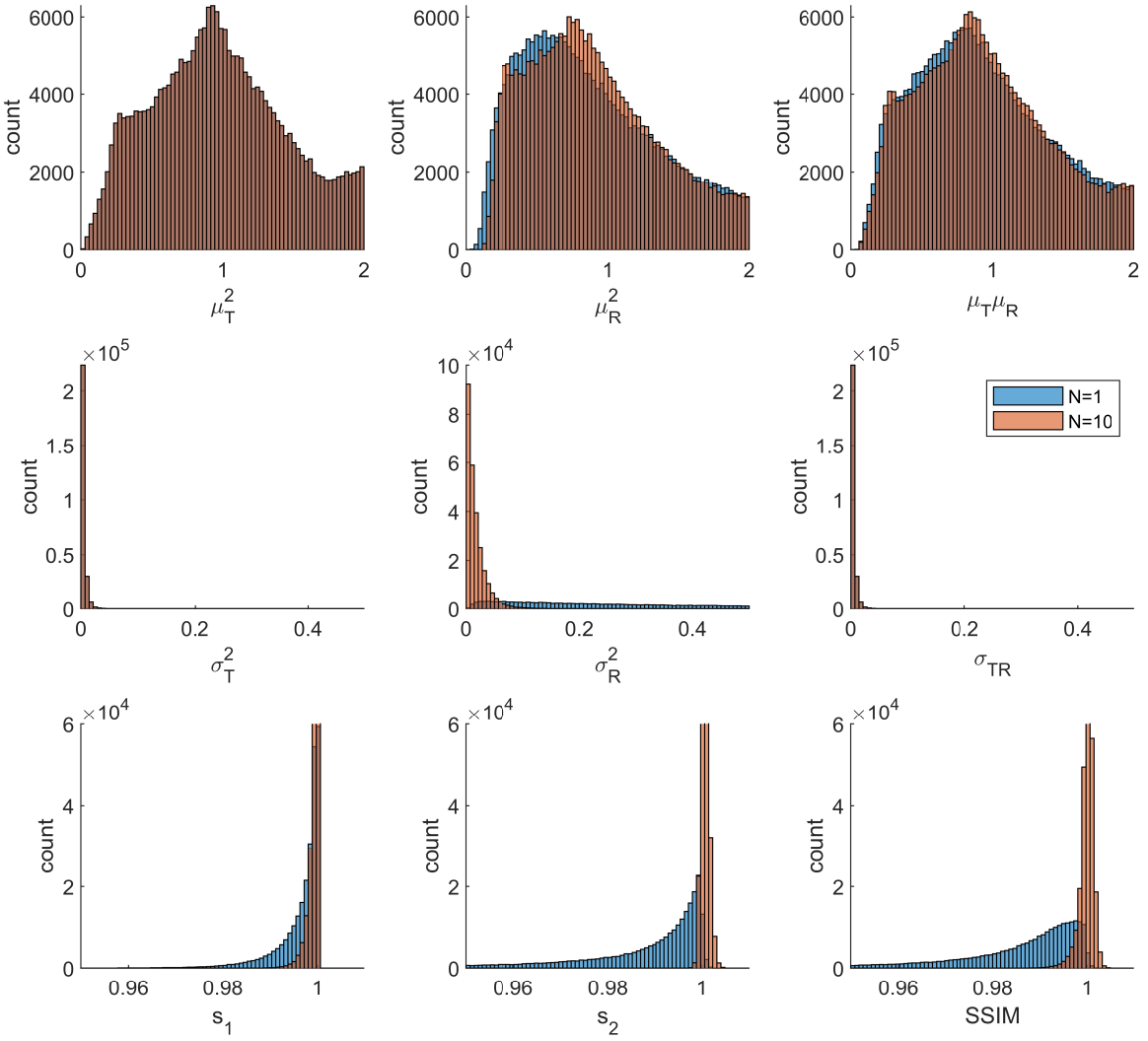


Fig. A.3 Evolution of components of the SSIM equation over time. Image used is the 512×512 *Mandrill* test image.

A.5 Conclusion

This chapter presented a numerical relationship for estimating the bias and error against number of sub-frames for OSPR. This was found to provide accurate estimations of expected MSE for given amplitude distribution with prediction error within 1%. This was then extended to cover the case of SSIM improvement against sub-frame. This analysis again provided accurate estimates of convergence.

Unsurprisingly, though variance converges to zero, MSE does not converge to zero and SSIM does not converge to unity. This is due to the non-linear square relationship between

amplitude and intensity. Despite this, the improvement in MSE and SSIM is significant and OSPR remains a viable algorithm for real-time holography.

Appendix B

Benchmarking the Gerchberg-Saxton Algorithm

Appendix B

Benchmarking the Gerchberg-Saxton Algorithm[1]

B.1 Introduction

This section provides an in-depth benchmark of the Gerchberg-Saxton (GS) algorithm discussed in Section 2.4.3, perhaps the most common computer-generated holography (CGH) algorithm for projection applications using phase modulating devices.

While many high performance algorithms are available[266–270], less information is available on the precise factors influencing performance. In order to support our ongoing research into high power laser projectors this appendix benchmarks the widely used GS algorithm. GS is widely used in relatively "smooth" problems such as display on multi-level SLMs [50]. As shall be shown later, GS fails to converge for binary SLMs.

This work begins by discussing the test setup and constraints and discusses individual algorithm components including quantisation, error metrics, starting points and floating point precision. It then continues to expand in detail on performance factors for GS and develop a heuristic relationship for hologram generation performance.

[1]The contents of this section are currently included in the following preprint:

Peter J. Christopher, George S. D. Gordon and Timothy D. Wilkinson “Benchmarking the Gerchberg-Saxton Algorithm", *arXiv*, 2020, arXiv: 2005.08623

B.2 Considerations

B.2.1 Hardware

While a number of hardware architectures are available for hologram generation [291], graphics processing units (GPUs) are the most used and most flexible. Originally targeted for video games, they are seeing increasing use in the scientific and financial sectors. Other works have discussed the relative benefits of alternative architectures such as field programmable gate arrays (FPGAs) or digital signal processing units (DSPs) in great depth. This work seeks, instead, merely to compare the algorithms used across the different device types. While some quantitative values for generation time will be given, this appendix is primarily focussed on relative performance of algorithms on GPU devices.

All the tests discussed here were run on a GTX 1080 GPU with additional system details presented in Table B.1. The processes were run on an independent machine with no additional workload and each test was run many times to ensure consistency and provide estimates of variance. By using low-level interfaces through C/C++, it was possible to manage potential memory bottlenecks to ensure fairer testing.

The GPU was accessed directly through the native Cuda interface to reduce potential issues with third party libraries. The only major library used is the Cuda FFT library (cuFFT) which is developed by the GPU manufacturer, Nvidia. Independent performance tests are discussed below.

Table B.1 Benchmarking hardware

Workstation	Details
GPU	GTX 1080, 2560 cores, 8GB DDR5 ram, 1607MHz clock
CPU	Intel® i7-7700K, overclocked to 4.5GHz, 4 cores, hyperthreaded to 8 cores
RAM	48GB 2400MHz DDR4
OS	Windows 10 Pro, Build 10.0.17134
HD	500GB M.2 SSD

B.2.2 Software

The benchmarking tests and output figures are generated by HoloGen, presented in greater detail in Chapter 4.

B.2.3 FFT Performance

The algorithm implementations discussed here are based on NVidia's cuFFT package. In an idealised world, FFTs can be generated in time $O(N^2 log(N^2))$. This is rarely achievable, however, due to the limitations and ever increasing complexity of real-world devices. While many benchmarks are available for performance, comparing this information between machines is non-trivial.

Traditionally FFT performance is significantly better when the resolution is a power of 2. A test of 500,000 FFTs with random resolutions between $2^1 1$ and $2^1 2$ factorised solely into primes larger than 100 concluded that the performance cost is of approximately $2.6 - 3.1$ times that expected of a resolution power of 2.

The performance for the 255 unfactorisable primes between $2^1 1$ and $2^1 2$ had performances significantly greater than 10 times worse. In order to reduce this factor, every sampling point is equal to a power of 2 multiplied by up to 2 values from the set $3, 5 and 7$.

Figure B.1 shows the behaviour of cuFFT against resolution along with the $O(N^2 log(N^2))$ trend line. The drop in the final measurement appears to be due to cuFFT making the transition to the 64 bit kernel.

The FFT calculations took up $> 98\%$ of the algorithm runtime of all the algorithms discussed here. Of the $> 100,000,000$ FFTs run during testing, $< 2\%$ of the runtime was spent on memory movements onto the GPU, transpose operations and quantisations. For analysis, therefore, the FFT time is assumed to dominate. Additionally, the shift, transpose and quantisation steps all naïvely run in $O(N^2)$ time.

A two dimensional FFT looks like the following:

$$\texttt{fft2} \rightarrow \texttt{transpose} \circ \texttt{fft} \circ \texttt{transpose} \circ \texttt{fft} \tag{B.1}$$

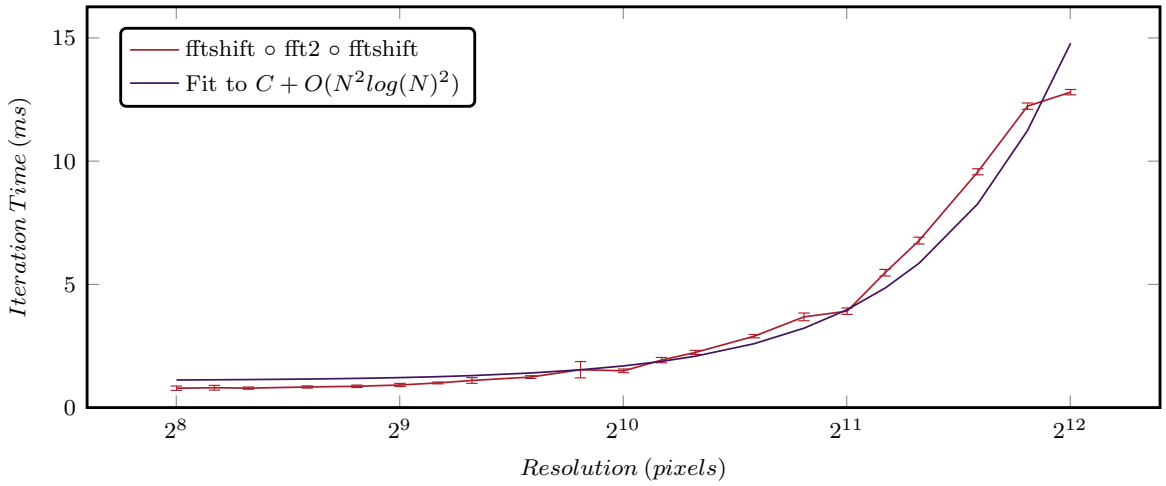


Fig. B.1 Comparison of FFT performance against image resolution using Nvidia cuFFT. Error bars show the 2σ confidence interval measured from 100 independent runs of 1000 pairs of FFTs and IFFTs

FFTs as used in holography require the zero order to be shifted to the centre of the image. This requires two additional `fftshift` operations

$$\texttt{fftshift} \circ \texttt{transpose} \circ \texttt{fft} \circ \texttt{transpose} \circ \texttt{fft} \circ \texttt{fftshift} \tag{B.2}$$

where `fftshift` is an operation to swap diagonal quadrants of an image.

B.2.4 Floating Point Precision

Traditional CPU arithmetical calculations are done using a math coprocessor or floating point unit (FPU). On modern Intel x86/x86-64 systems, the precision of the floating points used make little difference to calculation time and double (64-bit) precision is standard. On Nvidia graphics cards, single precision (32-bit) offers performance boosts of $2\times$ over double precision. Modern GPUs (SM_53 or later) are capable of working in half precision though the speed increase is $< 2\times$.

HoloGen is capable of being compiled in double, single and half precision variants. Running a suite of tests on the workstation machine described in Table B.1 on all three precisions showed that single precision was $\approx 100\%$ faster than double precision in almost all cases

while half precision only offered $\approx 30-50\%$ speed improvements. This is summarised in Table B.2.

The half precision performance in expected to improve with GPU generation and offers significant promise for hologram algorithms. The linearity of the FT means that values can be normalised near to 1, reducing the impact of reduced exponent bits. Incremental algorithms such as GS only run for $10-100$ cycles, reducing the impact of accumulated errors while longer algorithms like DS which can run to $> 100,000$ cycles don't introduce incremental errors.

The results shown here are generated for single precision as this is the standard in similar analysis.

Table B.2 Impact of floating point precision on hologram generation speeds

Precision	Time Modifier	Description
Single (32-bit)	1	1 sign bit, 8 exponent bits, 24 significand bits
Double(64-bit)	1.96-1.99	1 sign bit, 11 exponent bits, 53 significand bits
Half (16-bit)	0.67-0.76	1 sign bit, 5 exponent bits, 11 significand bits

B.2.5 Quantisation

Spatial Light Modulators (SLMs) are unable to arbitrarily modulate light, instead typically modulating only in discrete values of phase or amplitude. The act of modifying a target light field to meet these constraints is here referred to as *quantisation*.

By far the most dominant quantisation approach is the *nearest-neighbour* approach where the pixel is quantised to the nearest complex value achievable on the device. Other schemes such as *error diffusion* exist but can be considered algorithms in their own right and are beyond the scope of this appendix. As modern SLMs are usually discretely addressed, they present a number of modulation levels which typically vary between 2^1 and 2^8.

B.2.6 Test Setup

The majority of simulations run assumed an SLM twice the size of the target image. This is to compensate for the rotational symmetry inherent in binary holograms as shown in Figure 2.7. This corresponds to the format shown in Figure 2.7 (a). Different SLM modulation schemes offer different challenges in terms of conjugate symmetry and zero order behaviour.

B.2.7 Starting Point

For many algorithms, the starting point is often a significant factor in convergence time. Two main approaches are often used:

- **Randomisation** - random points in the unit circle are generated for each diffraction field pixel. These are then quantised to the SLM modulation behaviour.
- **Back Projection** - The target replay field is back projected to the diffraction field and quantised to the SLM modulation behaviour.

Often back projection is preferable for a single quick solution with randomisation being used for better quality, slower results. While other starting points exist, they are little used and can be considered beyond our scope.

B.2.8 Merit Function and Error

In this appendix error is exclusively treated as PI MSE $E = E_{MSE,PI}$ but it should be noted that other formulations are available as discussed in Section 2.4.1. Additionally, normalised MSE (NMSE) is introduced where the MSE of each test image is normalised to that of the benchmark *Mandrill* test image by the ratio of the convergent graphs. This allows for using a variety of test images while keeping the results simple to understand.

$$NMSE_{image,n} = MSE_{image,n} \frac{MSE_{Mandrill,1000}}{MSE_{image,1000}} \tag{B.3}$$

B.2.9 Convergence

Mathematically a series x_n is *convergent* to value L if for any given ε there exists a value n for which the following applies

$$|x_t - L| < \varepsilon \quad \forall \quad t \geq n \tag{B.4}$$

and *diverges* otherwise. This is not practical for use in a non-linear piecewise context.

Instead, for this application, the series NMSE is taken as converging to value L being the first series element n for which the following applies.

$$|NMSE - L| < \varepsilon \quad \forall \quad 2n \geq t \geq n \tag{B.5}$$

The value of ε is defined arbitrarily to be 0.001.

B.3 Gerchberg-Saxton Algorithm and Variants

This section discusses the first algorithm under consideration: Gerchberg-Saxton. Originally presented in 1972 as a means of phase retrieval,[50, 292] GS quickly became a common means of hologram generation[293, 51].

GS is part of the wider family of Iterative Fourier Transform Algorithms (IFTAs) which operate by transforming an image between the diffraction field and replay field while enforcing the constraints of both as shown in Figure 2.14a. It can be shown that the IFTA approach minimises MSE as well as producing identically independently distributed (i.i.d.) results for i.i.d. randomised phase inputs[294–296]. This only applies, however, in the unmodulated case. For modulated cases divergence is common.

Figure B.2 shows the convergence with 2σ confidence interval for the test images shown in Figure 2.12. Each line is taken as the average of 50 runs for each of the 6 standard images and 10 modulation levels over 30 iterations. A total of 3,000 independent runs. The simulated device is phase modulating and a randomised starting point is used. This can be compared with Figure B.3 where an IFFT of the target image is used, a so called back projected starting

point. It will be seen that convergence for GS is extremely fast, plateauing after only about 10 iterations.

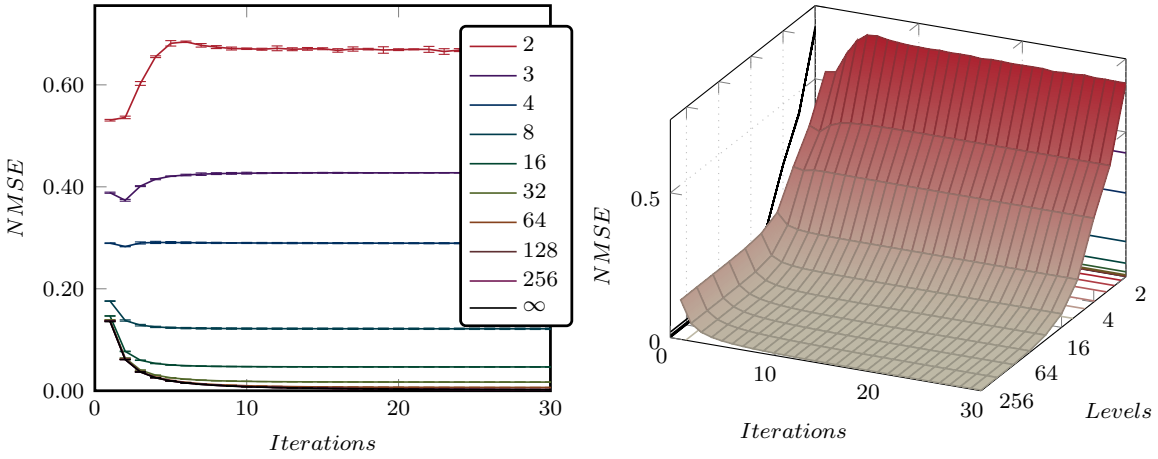


Fig. B.2 Gerchberg-Saxton convergence with randomised start for numbers of modulation levels. Error bars show the 2σ confidence interval. Each line is taken as the mean of 50 independent runs for each of the 6 standard test images shown in Figure 2.12.

For binary holograms, and other low level count holograms, cycles were often observed in the solution. Rather than converging to a single static result, a cycle of upwards of 10 elements would form. This periodic solution could have significant impact on the MSE with variations of up to 9%. For optimum results with GS, it should not be assumed that MSE always improves.

Another observation confirms that by Cable [73] that beyond a certain point , more modulation levels does not improve the quality of the hologram. A continuously modulated hologram offers less than a 1% improvement on a 64 level hologram. This can be seen in Figure B.4. Certainly, there is no need for greater than 8-bit addressed real-world systems (256 levels).

A log-log plot of convergent error (where convergence is taken as being given according to (B.5)) against number of modulation levels is shown in Figure B.4. Error bars show the 2σ confidence interval. Each line is taken as the mean of 50 independent runs for each of the 6 standard test images shown in Figure 2.12. The linear relationship is of interest and the authors are unaware of it having been reported previously. This result is not unexpected for the system, but it is good to see it confirmed here.

The behaviour against resolution is also of interest. Figure B.5 shows the error for a binary (left) and multi-level (right) device after set numbers of iterations for resolutions ranging between 128×128 up to 1024×1024. The flatness of the lines is of interest, suggesting

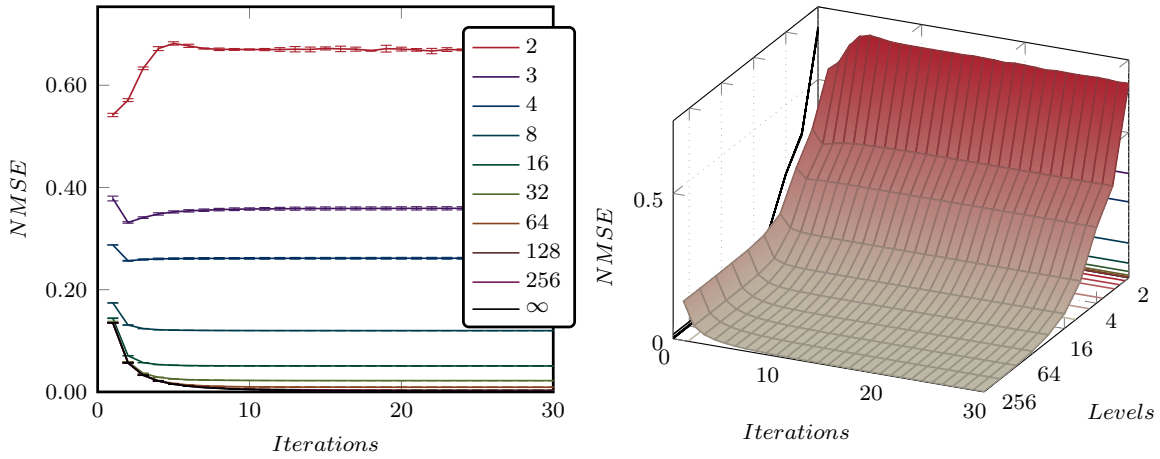


Fig. B.3 Gerchberg-Saxton convergence with back projected start for numbers of modulation levels. Error bars show the 2σ confidence interval. Each line is taken as the mean of 50 independent runs for each of the 6 standard test images shown in Figure 2.12.

the final pixel error and number of iterations until convergence are both independent of the resolution of the test image.

Finally, Figure B.6 shows the iteration time against resolution. The jump in the graph at $N_x = N_y \approxeq 2^{8.5} \approxeq 360$ is due to a similar jump in the calculation time in the cuFFT library. Excluding this jump, as expected for a cuFFT based process, the graph fits with $< 5\%$ MSE to the expected $O\left(N^2 \log(N)\right)$ curve.

B.3.1 Performance

While an analytical relationship of problem performance is impossible, an approximate formula can be used for estimating process performance for given a system and parameters. Using the > 100 million iterations calculated, the following approximate formula is suggested that may be used for estimating run time t for a given system.

$$t_{GS} = C_{numitr} C_{machine} C_{precision} C_{software} (C_{itr_1} + C_{itr_2} N_x N_y \, Log(N_x) Log(N_y)) \text{ms} \qquad \text{(B.6)}$$

where

- $C_{itr_1} \approx 0.71$ and represents the memory transfer and input/output component of the operation.

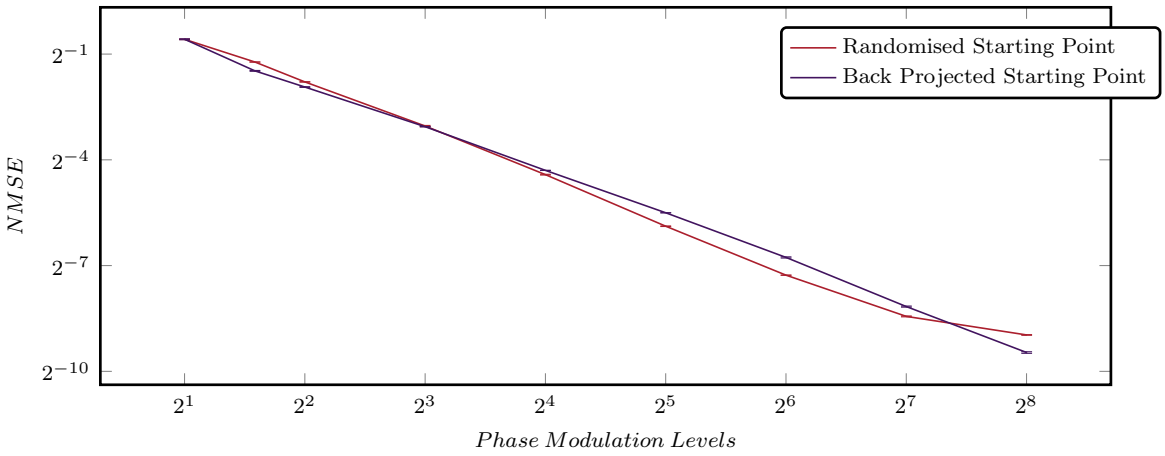


Fig. B.4 Gerchberg-Saxton convergent error against number of modulation levels for starting points.

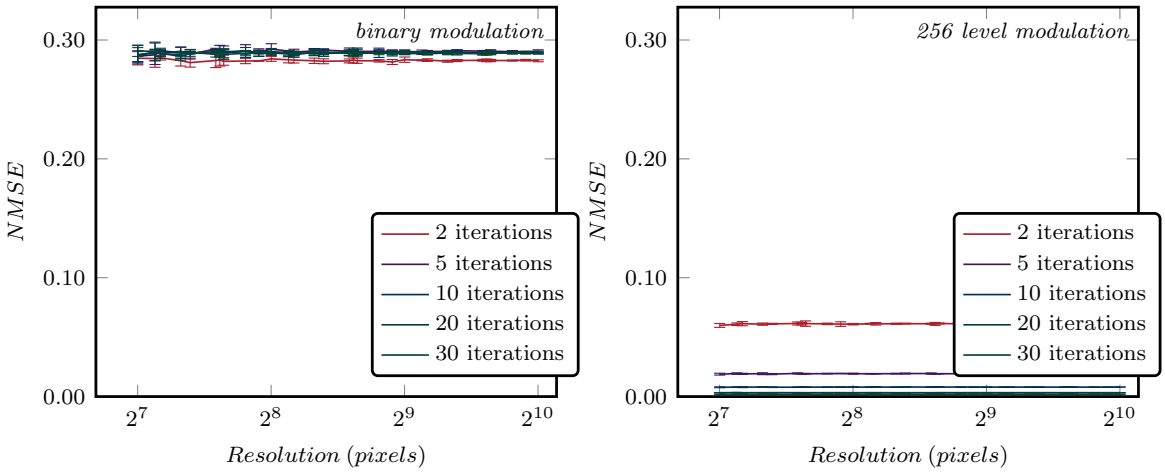


Fig. B.5 Gerchberg-Saxton error against resolution for binary (left) and 256 (right) level modulation.

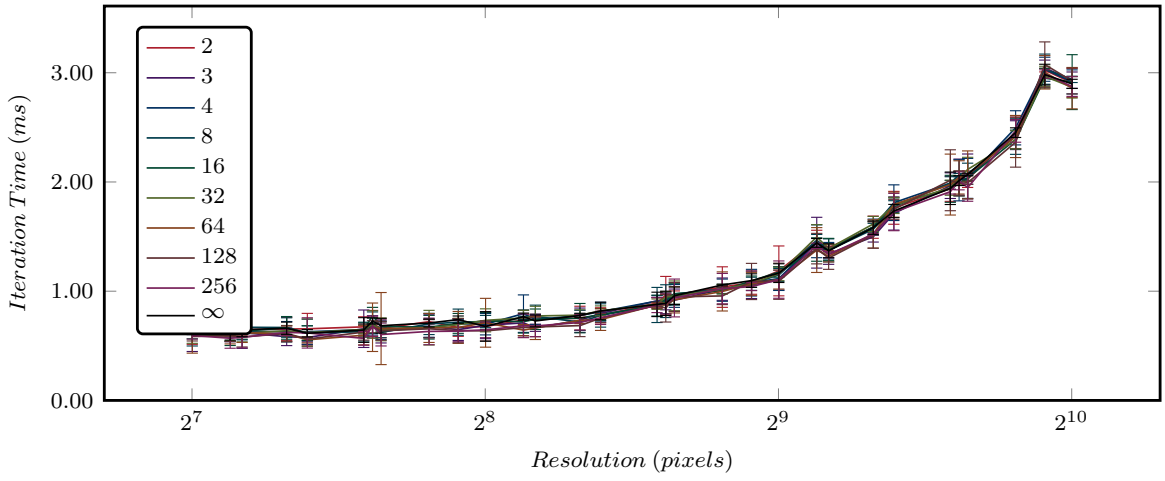


Fig. B.6 Gerchberg-Saxton iteration time against resolution for numbers of modulation levels.

- $C_{itr_2} \approx 1.09 \times 10^{-6}$ and represents a scaling constant
- $C_{machine}$ is a parameter of the machine used and is equal to 1 in this case. As the process is GPU bound, $C_{machine}$ can be treated as $C_{machine} \propto \frac{GFLOPS}{102.8}$ where 102.8×10^9 is the base FLOP rate for the GTX 1080 test device.
- $C_{software}$ is a parameter of the code developed and is equal to 1 in this case. As the software is applied to the general case and, it is estimated that this could be straightforwardly reduced to 0.5 for very specific cases.
- C_{numitr} is the number of iterations.
- $C_{precision}$ is a factor for floating point precision and is equal to 1 in this case. $C_{precision}$ can be determined from Table B.2.

While all of these parameters can be expected to vary from system to system, this heuristic relationship can be used as a starting point for system design. Verifying this relationship for all devices is impossible but tests were run on seven different workstations varying in purchase date from 2011 to 2019 and with rated GFLOP performances varying by more than two orders of magnitudes. All seven machines demonstrated performances within 20% of the relationship given here with the value of C_{itr_1} being the primary variable.

B.4 Conclusion

This appendix has benchmarked the Gerchberg-Saxton algorithm performance on a modern workstation. GS was found to offer high performance in the case of multi-level devices but to diverge in the case of binary devices. The number of iterations taken for convergence and the error after a given number of iterations was found not to depend on the target resolution. The starting point for the algorithm was also seen to have little bearing on the convergent error and convergence time. It was also seen that the benefits of increasing the number of modulation levels per pixel was seen to decrease according to a power relationship and it was suggested that the maximum number of modulation levels required by real-world systems is approximately 2^6.

A heuristic relationship for calculating the expected runtime of the algorithm depending on a wide array of system parameters was also presented. This allows for approximate

specification of device requirements as a function of the system, resolution and device floating operations per second (FLOPs).

Appendix C

Relative Limitations of Increasing the Number of Modulation Levels in Computer-Generated Holography

Appendix C

Relative Limitations of Increasing the Number of Modulation Levels in Computer-Generated Holography[1]

C.1 Introduction

As discussed in Section 1.3.2 most SLMs are capable of modulating either the amplitude or the phase in exclusion of the other. An appropriate choice of SLM for the application to hand is hence important, and the restrictions imposed by the SLM have to be accounted for when generating a suitable hologram for display. A clear advantage of binary SLMs is their switching speed, but for static holograms a user might expect a multi-level SLM to impose fewer constraints and to yield a field pattern closer to the desired target image. This paper aims to show that, while this intuition is true in the case of multi-phase SLMs, this is not the case for multi-amplitude SLMs that are found to not offer a marked improvement over binary-amplitude SLMs. Numerical and heuristic justifications of this are also presented.

[1]The contents of this section were previously included as part of the following publication:

> Peter J. Christopher and Timothy D. Wilkinson, “Relative limitations of increasing modulation levels in computer-generated holography", *Optics Communications*, (2020), Volume 462, DOI: 10.1016/j.optcom.2020.125353

C.2 Test Image

The test image used for this work is the *Mandrill* image from the USC-SIPI database[49] with a randomised phase profile. The conjugate symmetry of the Fourier transform necessitates that all on-axis amplitude holograms have 180° symmetry. To improve the comparison between phase and amplitude holograms, rotational symmetry is artificially added to the Mandrill test image as shown in Figure C.1 (left).

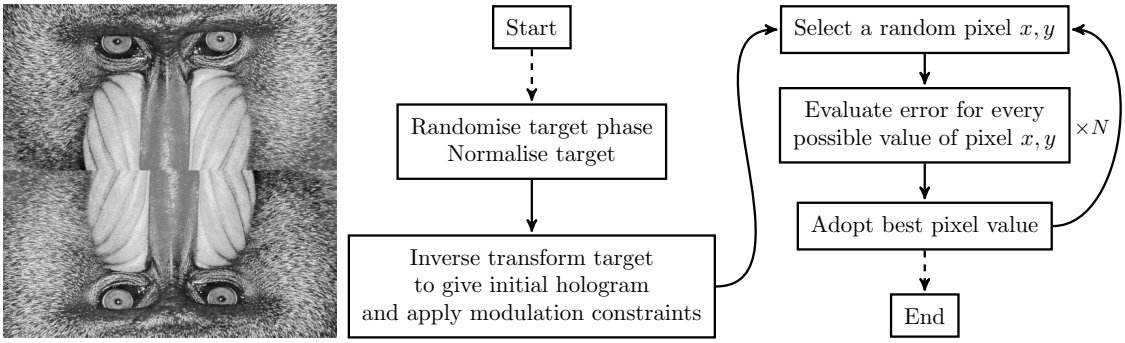


Fig. C.1 Mandrill test image magnitudes with artificial rotational symmetry (left) and flow chart detailing the hologram generation algorithm used (right)

C.3 Single Pixel Modification

In order to discuss the impact of additional modulation levels on the hologram reconstruction the effect of changing a single pixel independently is first discussed.

The inverse DFT of the Mandrill image of Figure C.1 (left) is taken and the magnitude of all obtained pixel values normalised in accordance with Parseval's theorem, corresponding to the initial modulation required for a phase-only SLM. The phase of a single randomly selected hologram pixel is altered by a random angle in the range $[-\pi, \pi]$ relative to this initial phase angle, and the resultant error ΔE is calculated for the replay field. Ten independent tests are shown in Figure C.2 (left). The plotted lines shown represent the errors achievable on a multi-level device for a given phase angle while the dots show the errors shown achievable on a binary device. It is observed that the binary phase modulation values rarely correspond to the lowest error achievable on a multi-level device. For our test image, a 2^8 multi-level phase SLM would offer a lower MSE when compared to a binary phase modulator more than 99% of the time.

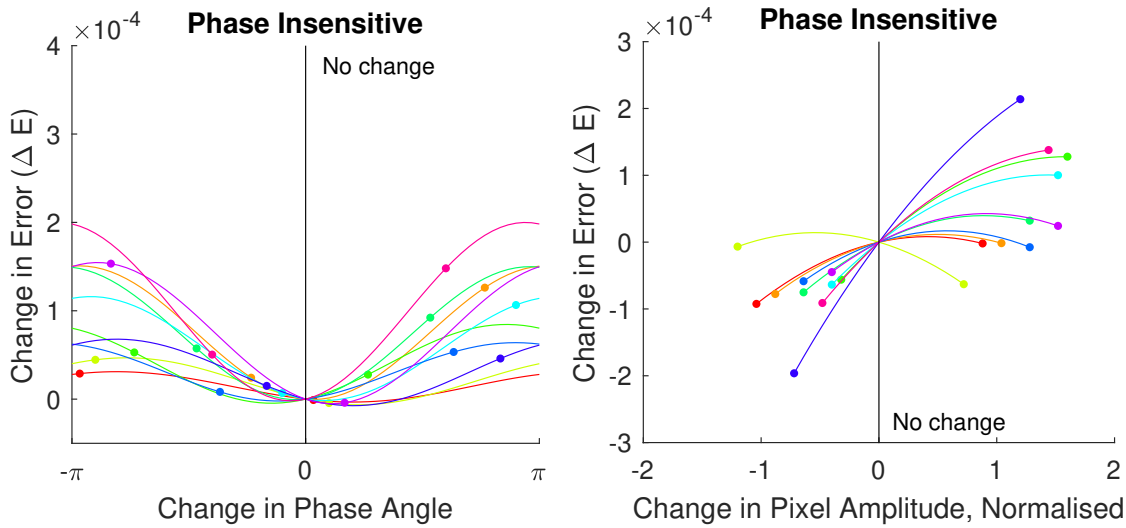


Fig. C.2 Effect of changing the value of a random selection of ten individual pixels on the MSE for phase (left) and amplitude (right) holograms. Dots show the values achievable on a binary device.

To investigate modulation of an amplitude-only SLM the inverse DFT of the Mandrill image of Figure C.1 is again taken, but in this case the phase-angle of each pixel is set to $0°$. The amplitude of ten randomly selected pixels is then varied in the range $(0,2)$ with the associated change in error ΔE calculated with results shown in Figure C.2 (right). In this case, it can be seen that the lowest error value often occurs at extremal values, corresponding to the values that can be obtained with an equivalent binary SLM. For our test image, a 2^8 amplitude level SLM would have offered better error improvements over a binary amplitude modulator less than 3% of the time.

The combination of these two results suggests that multi-phase modulation offers improvements over binary-phase modulation for almost all pixels individually while multi-amplitude modulation does not offer similar improvements over binary-amplitude modulation. To explore this further the case of generating an entire hologram is considered.

C.4 Algorithms

A wide array of hologram generation algorithms are available but many of these are limited to a specific class of problem. For example, Gerchberg-Saxton (GS) is commonly applied only to phase modulated holography [50]. In order to provide the fairest comparison of binary vs.

multi-level devices and amplitude modulating vs. phase modulating devices A form of two common search algorithms is adopted - direct search (DS) and simulated annealing (SA) - introduced in Sections 2.4.4 and 11.

A lightly modified algorithm shown in Figure C.1 (right) is adopted. This is similar to a DS algorithm, except that the pixel under consideration is set to the best possible quantised value, rather than merely considering the current value and a single randomly-selected alternative.[266] This offers a better comparison between binary- and multi-level holograms as the randomness inherent in DS algorithms is eliminated at the expense of higher computational requirements.

C.5 Effect on Convergence

In order to confirm the above result in the context of whole image manipulation 4×10^5 iterations of the algorithm are performed. This is shown in Figure C.1 (right) for binary- and multi-level quantisation on phase- and amplitude-modulating SLMs. The results of this analysis are reported in Figure C.3, which shows the MSE convergence for the phase-insensitive replay field, and Table C.1, which reports the final MSE obtained in each case.

Table C.1 Algorithm convergence - final error as percentage of initial error

Quantisation Levels	Amplitude Modulation	Phase Modulation
Binary	24.57%	22.76%
256-Level	21.06%	5.522%

It is observed that a 2^8 level phase hologram offers a final error less than 25% of the binary phase case. A 2^8 level amplitude level hologram only offers a final error equal more than 85% of the binary amplitude case. The final image qualities are shown in Figure C.4

C.6 Summary and Interpretation

This paper has discussed the relative merits of binary vs. multi-level quantisation in holography. Section C.3 showed that on a pixel by pixel basis, the most desirable pixel phase

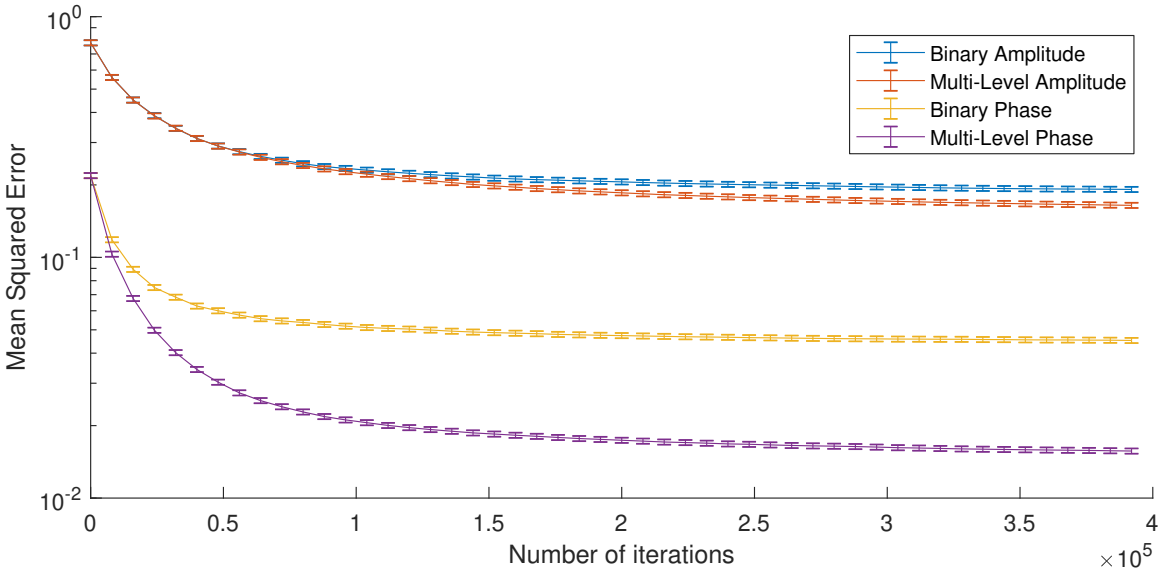
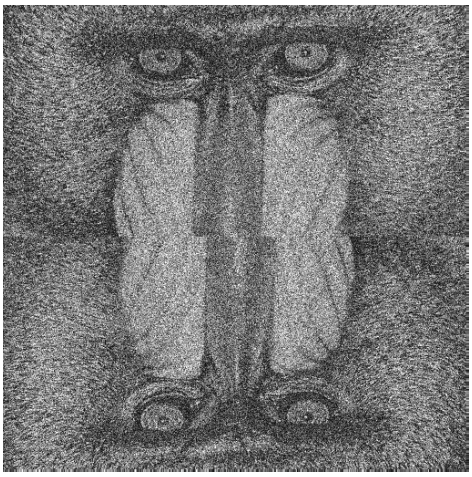


Fig. C.3 Comparison of convergence for multi-level and binary quantisation on phase-modulating and amplitude-modulating devices for the phase-insensitive replay field. Target is the *Mandrill* test image shown in Figure C.1 (left) with artificial rotational symmetry and a randomised phase profile. Values are taken as being the mean of 20 runs with independent random phase profiles with error bars showing two standard deviations. Error bars are shown for every 10,000th iteration to reduce visual clutter.

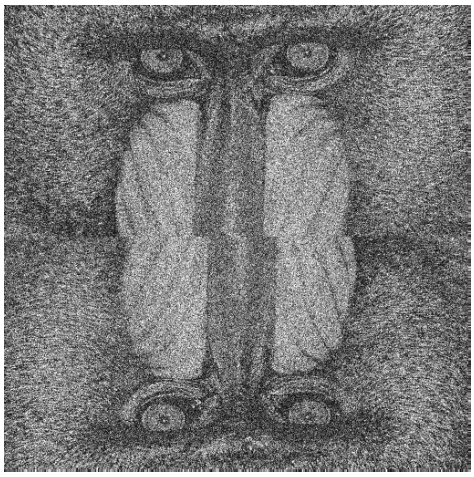
rarely corresponds to that achievable on a phase device. By contrast, for amplitude devices this hzsowed that on a pixel by pixel basis, the most desirable pixel value on a 256-level device is equal to the minimum or maximum value more than 85% of the time.

To explore this further, Section C.5 took a modification of the DS algorithm to explore the convergence of the two algorithms over time. As expected, both amplitude and phase holograms converged to a hologram with non-zero reconstruction error. The surprising element is that while multi-phase holograms massively outperformed binary phase holograms, the same could not be said for multi-amplitude holograms which only slightly improved on the binary-amplitude case.

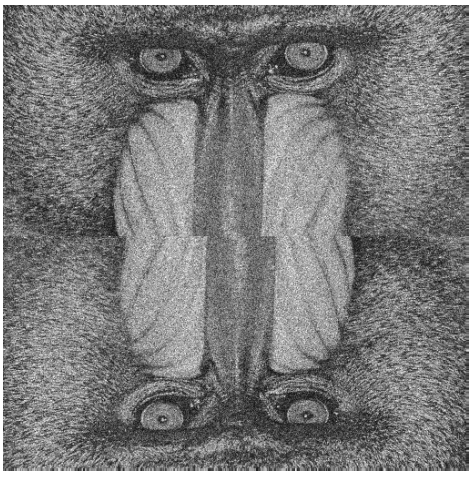
The authors believe that this result is due to the differences in modulation achievable between the regimes. A phase device does not change the amplitude of every spatial frequency, merely determining its location in the replay field. This leads to behaviour similar to that shown in Figure C.2 (left) where changing a single phase value has a sinusoidal impact on reconstruction error. An amplitude device, Figure C.2 (right), has a more complicated relationship as the angle of any one spatial frequency is fixed and it is the proportion that is variable. It can be imagined that for much of the time the actual preferred amplitude for a



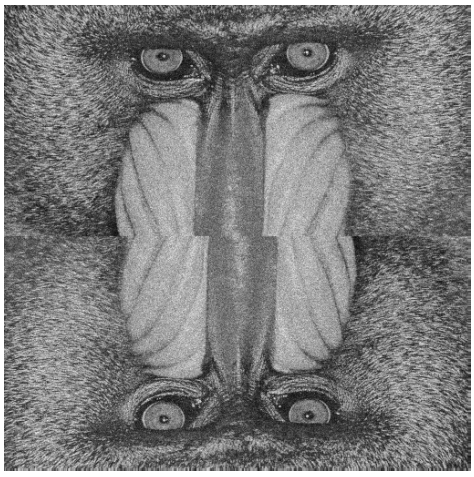
Binary Amplitude, MSE=0.209



Multi-Level Amplitude, MSE=0.205



Binary Phase, MSE=0.0509



Multi-Level Phase, MSE=0.0161

Fig. C.4 Converged images corresponding to Figure C.3. Target is the *Mandrill* test image shown in Figure C.1 (left) with artificial rotational symmetry and a randomised phase profile.

pixel would be either negative or greater than that achievable. In both cases, this would lead to the multi-level pixel being equal to the binary case.

C.7 Discussion

The analysis here has offered evidence of an observation worthy of further exploration. In particular, it is worth listing some of the inherent limitations in the work, however:

Firstly, the study has been entirely numerical in nature and does not take account of real-world issues such as lens aberration, non-flatness or speckle. Multi-level amplitude devices would offer a greater degree of flexibility when compensating for these issues.

Secondly, the error metric used was MSE due to its simplicity of use. Display applications often use the structural simularity index (SSIM) as a metric of visual quality due to the greater range of quality issues highlighted. It has recently been argued that the distinction between SSIM and MSE is not as significant as previously thought [287] and future investigation should explore the effect of modulation on SSIM and other error metrics.

Thirdly, the choice of algorithm deserves consideration. Many algorithms only function for a small number of different configurations. The search based algorithm used, Figure C.1, was chosen as it was equally applicable to every configuration considered [297, 298].

Fourthly, local minima are an expected issue with search algorithms. The authors suggest that any effect of this will disproportionately effect binary devices and that the difference in convergence shown in Figure C.3 is likely to overestimate the difference in *best possible* hologram for binary and multi-level cases. While impossible to quantify exactly, the influence of local minima on the convergence graph can be estimated from the magnitude of the standard deviation in the independent runs.

Fifthly, this analysis does not take account of more advanced optical configurations. A wide array of systems such as amplitude hologram encoding [299] have been developed. These use amplitude SLMs in clever configurations to get a greater degree of control in the reconstruction. The analysis here applies only to a simple single-SLM hologram in either a 2f or far-field configuration and while it is expected that the results will be more widely applicable, this has not been explicitly researched. For example, phase holograms can emulate some of the behaviours of amplitude holograms [300].

Sixthly, this work takes no specific account of the Fresnel regime, focussing exclusively on the Fraunhofer regime. While many of the assumptions made are equally applicable, this is worthy of further investigation.

Seventhly, only a single test image, the *Mandrill*, was considered. Independent tests with the *Peppers* test image produced similar results but the effect of target amplitude profile on this result is an area requiring greater exploration.

Eighthly, this analysis does not draw a conclusion on the relative merits of amplitude vs. phase holography, merely the effect of number of modulation levels. Other factors greatly influence that decision, for example phase SLMs are subject to phase compression and therefore require greater environmental control than their amplitude counterparts. Figure C.3 fails to take this and many other effects into account.

Ninthly, it is not possible to draw a conclusion on the relative merits of amplitude vs. phase holography, merely the effect of number of modulation levels. Other factors greatly influence that decision, for example phase SLMs are subject to phase compression and therefore require greater environmental control than their amplitude counterparts. Figure C.3 fails to take this and many other effects into account.

Tenthly, this letter does not include a discussion of other concerns in holography other than that of final error. For example, diffraction efficiency considerations factor into amplitude hologram design [301].

Eleventhly, only considered the case of a phase insensitive replay fields has been considered.

Twelthly, this work does not discuss any methods based on phase-only devices with a phase-shift larger than 2π.

Finally, the authors would be interested in this analysis's applicability to similar problems. For example, our approach might be applicable to a comparison of binary and continuous zone plates.

C.8 Conclusion

The impact of altering a single hologram pixel and of optimising an entire hologram have both been investigated for phase-only and amplitude-only holograms, with overall improvement

metrics provided in Table. C.1. It has been shown that, for the Mandrill replay field of Figure C.1 (left) generated using a variant of direct search, optimising the hologram for display on a 256-level phase SLM provides a significant improvement over optimising the hologram for display on a binary phase SLM. On the other hand, optimising the same hologram for display on a 256-level amplitude SLM only offers a marginal improvement over a binary amplitude SLM in both use cases.

A detailed discussion of the assumptions and limitations of this observation has been presented and it is hoped that this could help inform the decision making process during experimental design.

Appendix D

Lookup Tables for Phase Randomisation in Hardware Generated Holograms

Appendix D

Lookup Tables for Phase Randomisation in Hardware Generated Holograms[1]

D.1 Introduction

This paper seeks to present a method of improving the speed of hologram generation for two common hologram generation algorithms - Gerchberg-Saxton (GS) and One-Step Phase-Retrieval (OSPR) - in the context of embedded devices. This is done by moving elements of the computation dependent on a random number generator to a look-up table (LUT). It is shown that this can remove the need for trigonometric functions and a pseudo random number generator (PRNG) from the embedded process at the expense of higher memory usase. Significant reuse of the random data can occur in each frame before significant error is introduced. This is then demonstrated on an implementation of OSPR for a 1024×1024 binary ferroelectric display[41].

[1]The contents of this section are currently included as part of the following preprint:

> Peter J. Christopher, Ralf Mouthaan and Timothy D. Wilkinson “Lookup tables for phase randomisation in hardware generated holograms", 2020, arXiv: 2004.04049

D.2 Motivation

Implementations of CGH algorithms are able to use PRNGs[302] and trigonometric functions to carry out the phase randomisation. Field Programmable Gate Arrays (FPGAs) or Digital Signal Processors (DSPs) are also capable of this but can require the use of proprietary IP cores for the PRNG and trigonometric functions. To improve performance, Lookup Tables (LUTs) can be used. random numbers can be generated at compile time to fill the LUT which then acts as a *pool* of random numbers for algorithms to sequentially draw from. It the length of the random number LUT is given by N_{LUT} then the case where N_{LUT} is greater than the size of the test image is indistinguishable from the case without a LUT. LUTs of this size can be impractical on low-cost devices[89].

Figure D.1 shows an example of phase randomisation using LUTs. The *pool* of random numbers available from the LUT is cycled through sequentially with successive frames and sub-frames continuing from where the previous left off.

This appendix presents a heuristic approach for investigating the effect of N_{LUT} on edge enhancement and sub-frame independence. This is then tested on a binary phase ferroelectric device.

D.3 Hard limits

For the sub-frame independence and edge enhancement reduction motivations, hard limits exist for theoretical correspondence between the LUT and psuedo-random number generator approaches.

The hard limit for N_{LUT} to meet the sub-frame independence constraint is that N_{LUT} must be greater than N_{SF}, where N_{SF} is the number of OSPR sub-frames.

To meet the edge enhancement constraint for both OSPR and GS, to ensure the hardware LUT approach is equivalent to the software pseudo-random number generator case, N_{LUT} must be at least equal to $N_x N_y N_{\mathrm{SF}}$ where N_x and N_y are the x and y resolutions respectively. This is impractical on many hardware devices so instead it is attempted to determine the relationship between phase randomisation LUT length and final error.

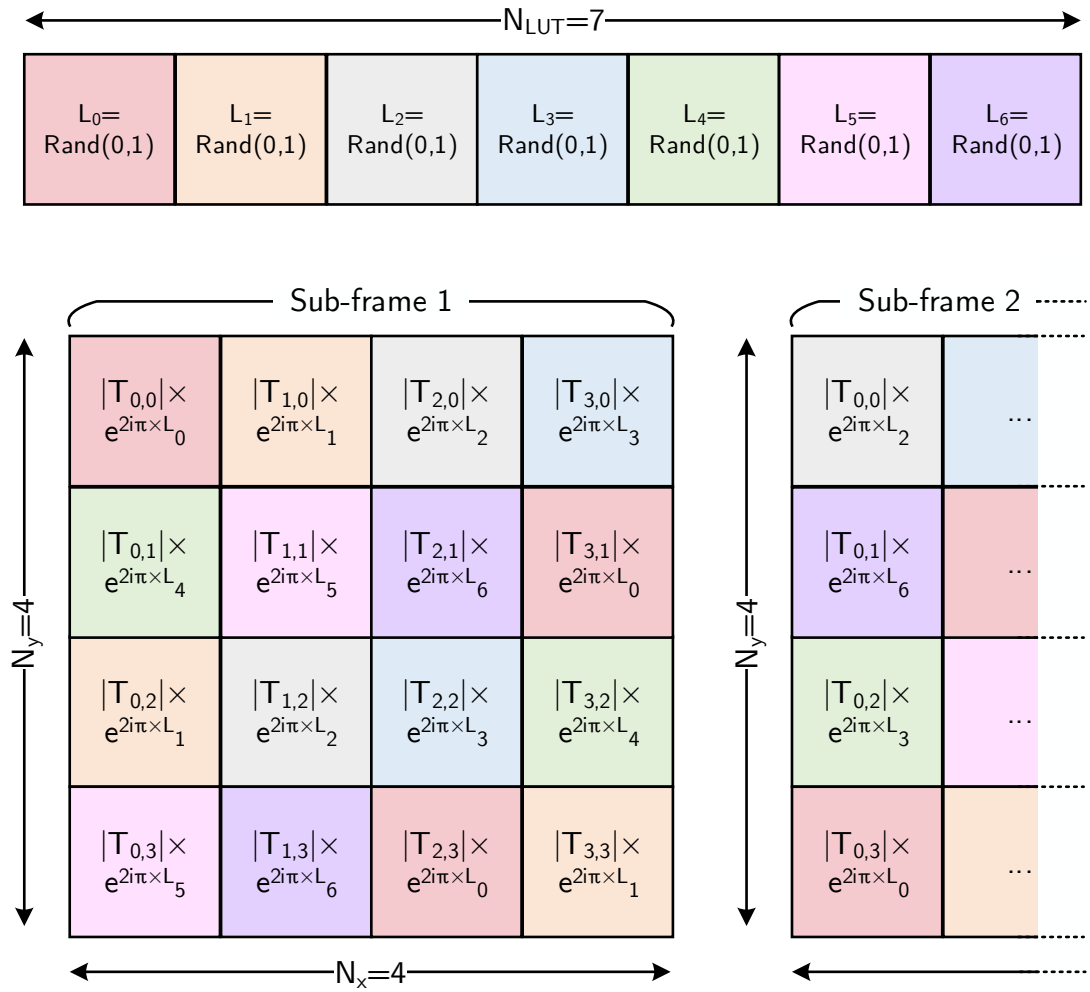


Fig. D.1 Example of phase randomisation using look-up tables, each block represents a single pixel.

D.4 Results

Figure D.2 shows the time averaged errors for prime values of N_{LUT} for OSPR binary phase holograms with 24 sub-frames for the six 256×256 pixel test images shown in Figure 2.12. Each value is the average of 100 independent runs for each of the 6 test images. A LUT for this would require over 12 Mb of space and realtime generation on a 60 FPS device would require multiple GFLOPs of processing power. Prime values of N_{LUT} are used in order to avoid factorisation issues. Each error value shown is the average of 100 independent runs and is normalised to the mean error for the *Mandrill* test image to give the Normalised MSE (NMSE).

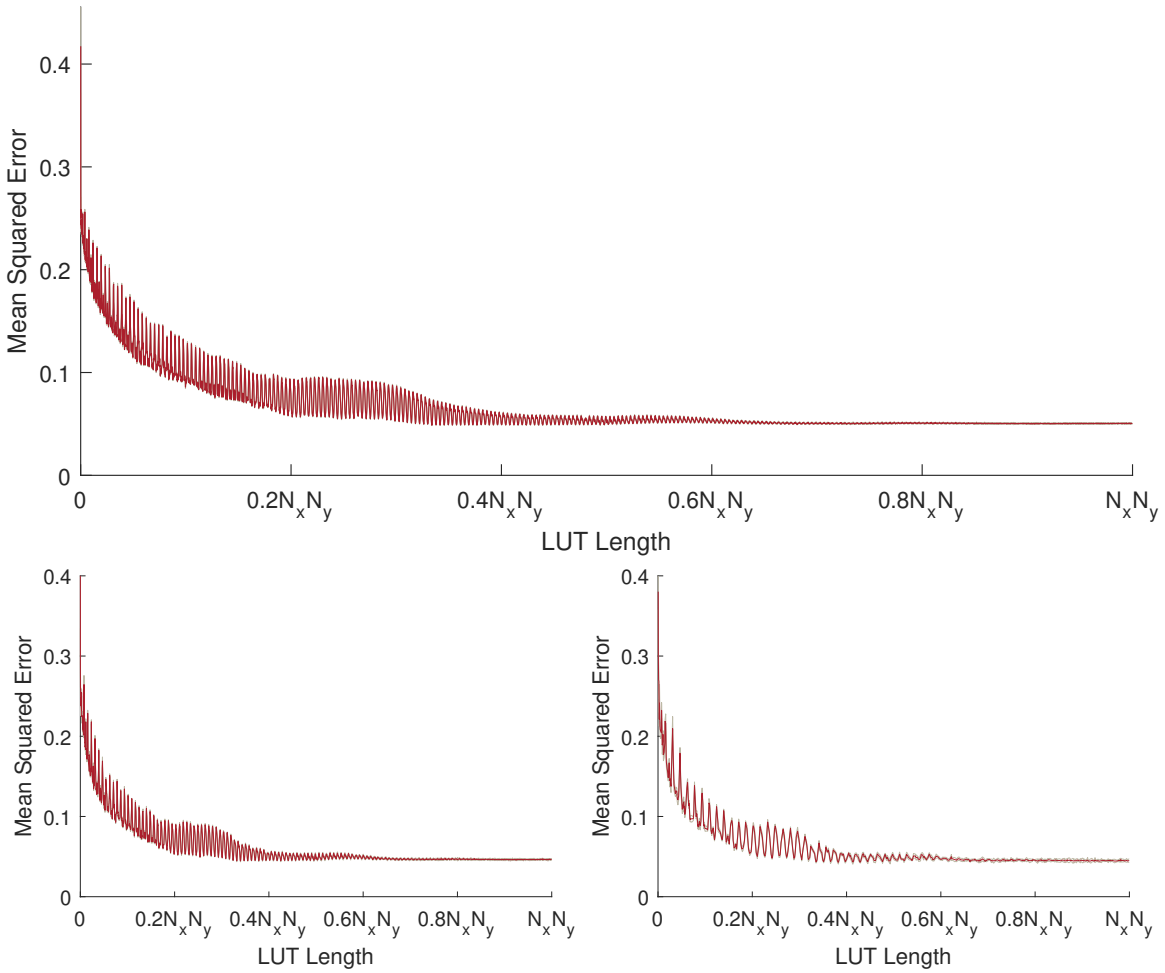


Fig. D.2 Time averaged errors for prime values of N_{LUT} for OSPR binary phase holograms with 24 sub-frames for the 6 256×256 (top) pixel test images shown in Figure 2.12. The mean value is shown in red with two standard deviations above and below shown in grey. A comparison with 128×128 (bottom left) and 64×64 (bottom right) holograms is shown for comparison.

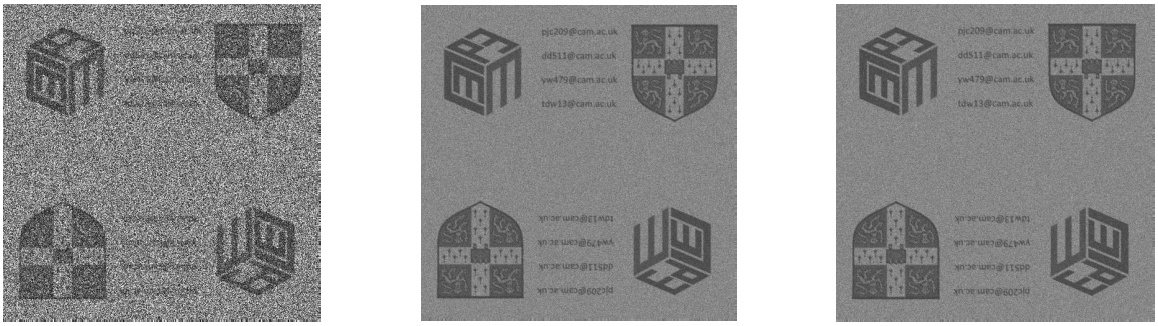


Fig. D.3 Single OSPR sub-frame (left) with time average of 24 OSPR sub-frames with independent phase randomisation (centre) and $N_{\mathrm{LUT}} = 10007$ (right).

$$NMSE_{image,n} = MSE_{image,n} \frac{MSE_{Mandrill,1000}}{MSE_{image,1000}} \tag{D.1}$$

The results show that, as expected, error is very high when N_{LUT} is 0 and drops off rapidly until $N_{\mathrm{LUT}} > N_{\mathrm{SF}}$. Descent is slower there after with significant variation. Spikes occur at regular intervals when the LUT would be close in length to a multiple of the resolution. For example $N_{\mathrm{LUT}} = 256$ leads to a sharp increase in error as every image row has the same phase behaviour.

Figure D.2 also shows the effect of varying the hologram size $N_x \times N_y$. The similarity between the graphs suggests that for a given target error the required LUT size will be proportional to the total hologram size.

Figure D.3 shows this effect on a real image. A single OSPR frame is shown right with 24 sub-frames in the centre. Both of these are with independent random numbers. The right hand image shown the case of $N_{\mathrm{LUT}} = 10007$, less than 0.03% of the naíve length, and shows less than 5% additional error. The value of 10007 for N_{LUT} was chosen as the first prime number larger than 10000 and similar results are seen for other large prime numbers. Independent trails with different test images showed no appreciable change in observed error. In practice the system admits of significant tuning for different test image resolutions as the oscillation in observed error is high in Figure D.2.

D.5 Validation

In order to validate these findings, the algorithm is demonstrated for a 1024×1024 binary ferroelectric display as shown in Figure D.4. The left hand image shows a fully independent randomisation while the right shows the case of $N_{\mathrm{LUT}} = 10007$, less than 0.03% of the naíve length. The value of 10007 for N_{LUT} was chosen as the first prime number larger than 10000 and similar results are seen for other large prime numbers. Independent trails with different test images showed no appreciable change in observed error.

Fig. D.4 Top portion of the replay field for 24 OSPR time-averaged sub-frames with independent phase randomisation (left) and $N_{\mathrm{LUT}} = 10007$ (right). Captured using a Canon 5D Mark III with a 24-105mm lens and a $1/60$ second exposure.

D.6 Performance Discussion

This chapter has presented a simple method for removing the requirement for PRNGs and trigonometric functions from FPGA implementations of CGH. The process has the advantage of requiring only a single look-up operation. As the LUT is read through cyclically, the memory location can be predetermined, improving performance at runtime. This performance benefit comes at the cost of reduced runtime flexibility and a larger LUT size.

For comparison, a fast PRNG approach such as M-sequences would offer a small algebraic overhead and still require the use of LUTs for the sine and cosine terms.

For systems requiring runtime flexibility, PRNGs with LUTs for sin and cosine functions are expected to remain the dominant technique. For specific applications known at compile-time it is expected that this technique could offer performance and complexity benefits over existing approaches.

D.7 Conclusion

This chapted initially reintroduced two reasons for phase randomisation in holographic projection: sub-frame independence and edge enhancement reduction. These have been discussed in the context of two algorithms: GS and OSPR. For embedded systems devices including FPGA and DSP these phase randomisation requirements can present a significant processing challenge and the use of LUTs can require many Mbs of storage for even small images

As a result of this it has been proposed that a LUT can be used and three primary constraints on the minimum length of the LUT were put forward. First that the LUT have a prime length to reduce the chance of matching the periodicity of the image. Second that the LUT have a length greater than the number of sub-frames and third that the LUT should be greater than the largest dimension of the image. For the case of using a Gerchberg-Saxton (GS) algorithm, only the first and third constraints apply.

The proposed method was demonstrated using a LUT of length 10007 for a simulated cases, less than 0.03% of the length required for complete sub-frame independence. These both conformed to our expectations with the simulated image showing little less than 5% additional error when using the LUT. Finally, this was demonstrated experimentally by generating frames for a binary phase projector. This also showed the expected behaviour with no visible quality differences when using a LUT of length 10007.

This result is of significance to embedded holographic systems as it negates the necessity for embedded complex number rotation operations and reduces the required computation and the overall implementation complexity. This comes at the cost of system flexibility.

Appendix E

SLM Characterisation

Appendix E

SLM Characterisation[1]

E.1 Introduction

The CMMPE group has access to a large number of SLMs or Holographic Display Panels. These are HDP-1280-2 'BlueJay' ferroelectric displays capable of a 2,500 FPS display rate. Originally developed by Displaytech, Inc on behalf of Light Blue Optics, Inc for use in their *Light Touch* projector they measure approximately 11×25 mm^2 with a resolution of 1280*1280 pixels[303]. A large number of these HDPs were made available after the termination of the *Light Touch* line. Table E.1 shows the device specifications. (HDPs)

The original aim of this thesis was to use arrays of these HDPS as micro-projectors which together could provide the necessary power capacity for sintering of metal powders. As planned, the system described in this chapter would have been capable of characterising individual HDP panels while in operation. A series of unavoidable delays, however, meant that development of the HDP drivers is still not completed, more than two years after the original deadline and this chapter focuses purely on the laser induced damage threshold (LIDT) tests for the panels that could be done while inert[2]. The state of driver development is discussed in more detail in [304].

[1]The contents of this section are currently included as part of the following preprint:

> Peter J. Christopher, Nadeem Gabbani, William O'Neill and Timothy D. Wilkinson "Automated laser induced damage threshold testing applied to a ferroelectric spatial light modulator", 2020, arXiv: 2004.00712

[2]The unused mathematical formulation for in-situ characterisation is the subject of a forthcoming paper.

Table E.1 HDP-1280-2 specification[303, 305]

Parameter	Specification
Package Size	4.7mm by 11.3mm by 24.8mm
Active Area	7.27mm by 7.27mm
Pixel Pitch	5.68μm by 5.68μm
Fill Factor	88.1%
Resolution	1280 pixels by 1280 pixels
Framerate	$\leq 2,500/\,\text{sec}$
Modulation Type	Binary Phase
DC Balance Period	100ms

E.2 Background

Laser induced damage thresholds (LIDTs) are an internationally standardised way of quantifying the threshold laser fluence required to cause damage in optical elements. This section presents an automated system for LIDT testing in accordance with the ISO 11254 and ISO 21254 standards[306, 307].

As part of ongoing research into the power handling capabilities of Spatial Light Modulators (SLMs) an automated system for measurement of damage threshold values is presented. This is demonstrated for a commercial liquid crystal on silicon (LCoS) device under continuous wave (CW) laser illumination. This substrate was chosen as a significant quantity of devices were available along with detailed accompanying power handling measurements for comparison.

For CW power sources, the optic is exposed at 10 locations to a laser of known beam diameter and power. The result is then examined under a high magnification optical microscope for visible damage. The laser power is varied between measurements with the LIDT being taken as the highest laser power for which damage is not observed on any of the 10 exposure sites. 'Damage' is here defined according to the ISO definition as any detectable change in the substrate.

As bulk heating is assumed to be the primary mechanism for damage under CW exposure,[308] LIDTs are often quoted with the associated beam diameter.[309, 310] The LIDT is given herein terms of power per area or W/cm^2 for a $1/e^2$ beam diameter $\varnothing$ given in μm or mm.

LIDTs may also be given in terms of the effective area equal to the ratio of laser power to maximum power density.[306, 311]

The ISO standards do not require a specific beam profile and only maximum beam intensity and beam diameter are required. In the case discussed here detailed manufacturer specifications were available for the beam diameter. For Gaussian illumination, the peak power is approximately $2\times$ the equivalent power of an equivalent uniformly distributed beam.[308]

The primary motivation for this work is the automation of a task for improvement in speed and reduction in human error and begins by presenting the experimental setup and automation arrangements with a focus of methodology. The system is then validated using an LCoS device as a test case. Finally, the measured response is discussed and conclusions are drawn.

E.3 Experimental Setup

Manual LIDT measurement is straightforward requiring only a known light source, a means of attenuation, a substrate and a microscope. A 'plug-and-play' automated system requires little further work. Some devices such as polarising filters require light of known polarisation and the damage thresholds for components can range significantly.

Figure E.1 shows the schematic of the system designed for automating this process. The Computer Aided Design (CAD) design is shown in Figure E.2 along with its real-world implementation. A 'plug-and-play' approach is taken for the laser source which can include a range of directly cage mountable sources including diodes as well as fibre launched light sources.

The laser beam - red in Figure E.1 - is passed vertically downwards through a window and adjustable linear polariser. Adjustment of the polariser relative to the fast axis of the polarising beam splitter (PBS) allows for intensity control in elliptical beams and ensures a know polarisation on the sample. A 90:10 or higher beam splitter extracts a portion of the power for power measurement and a switchable neutral density (ND) filter ensures compatibility with a wide range of intensities. Len 4 acts as a telescope with the distance between the lens and the stage defining the incident beam spot. Integration with Zemax OpticStudio allows the control system to automate this process. Any reflected light is captured by the beam dump.

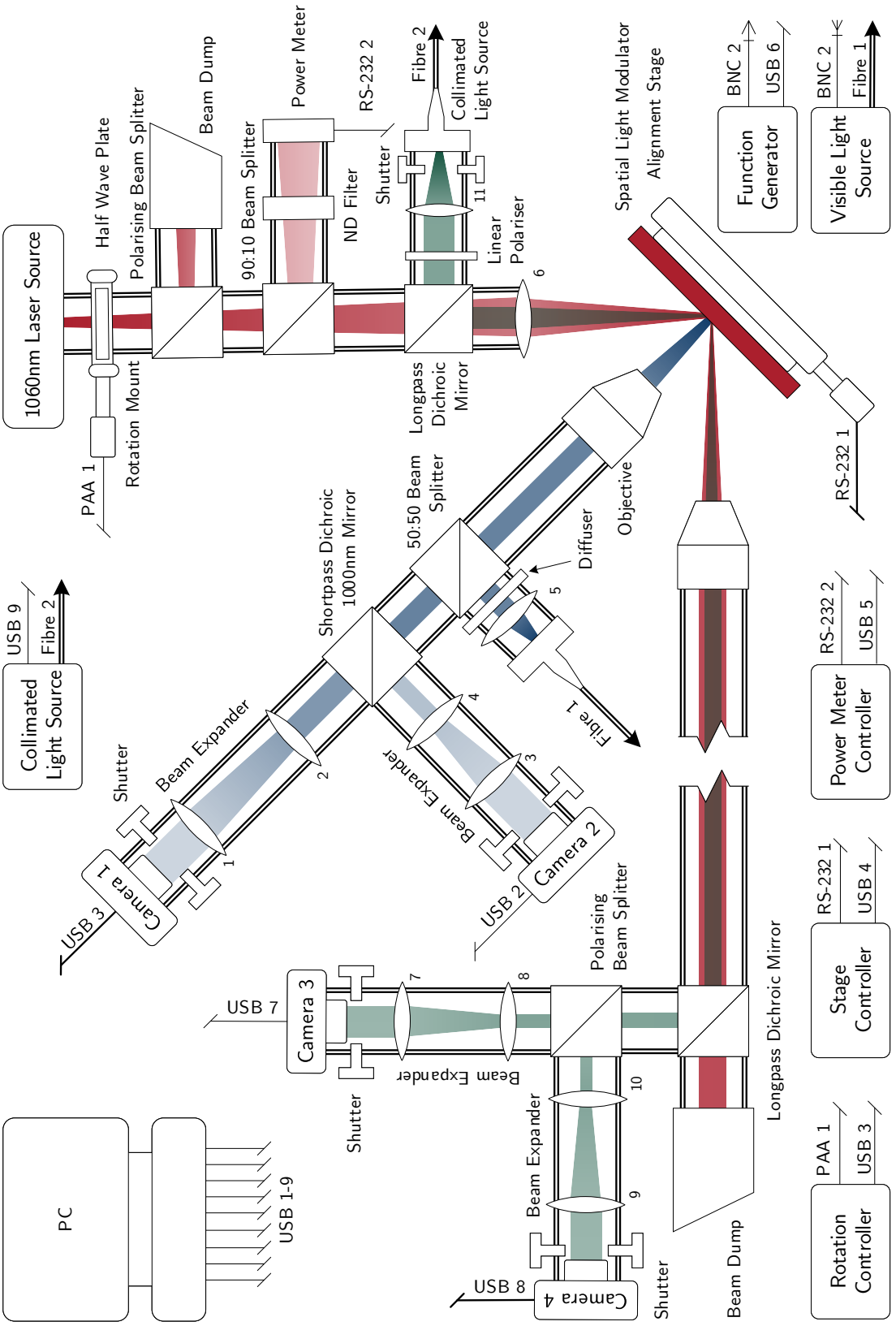


Fig. E.1 SLM power capacity experiment schematic

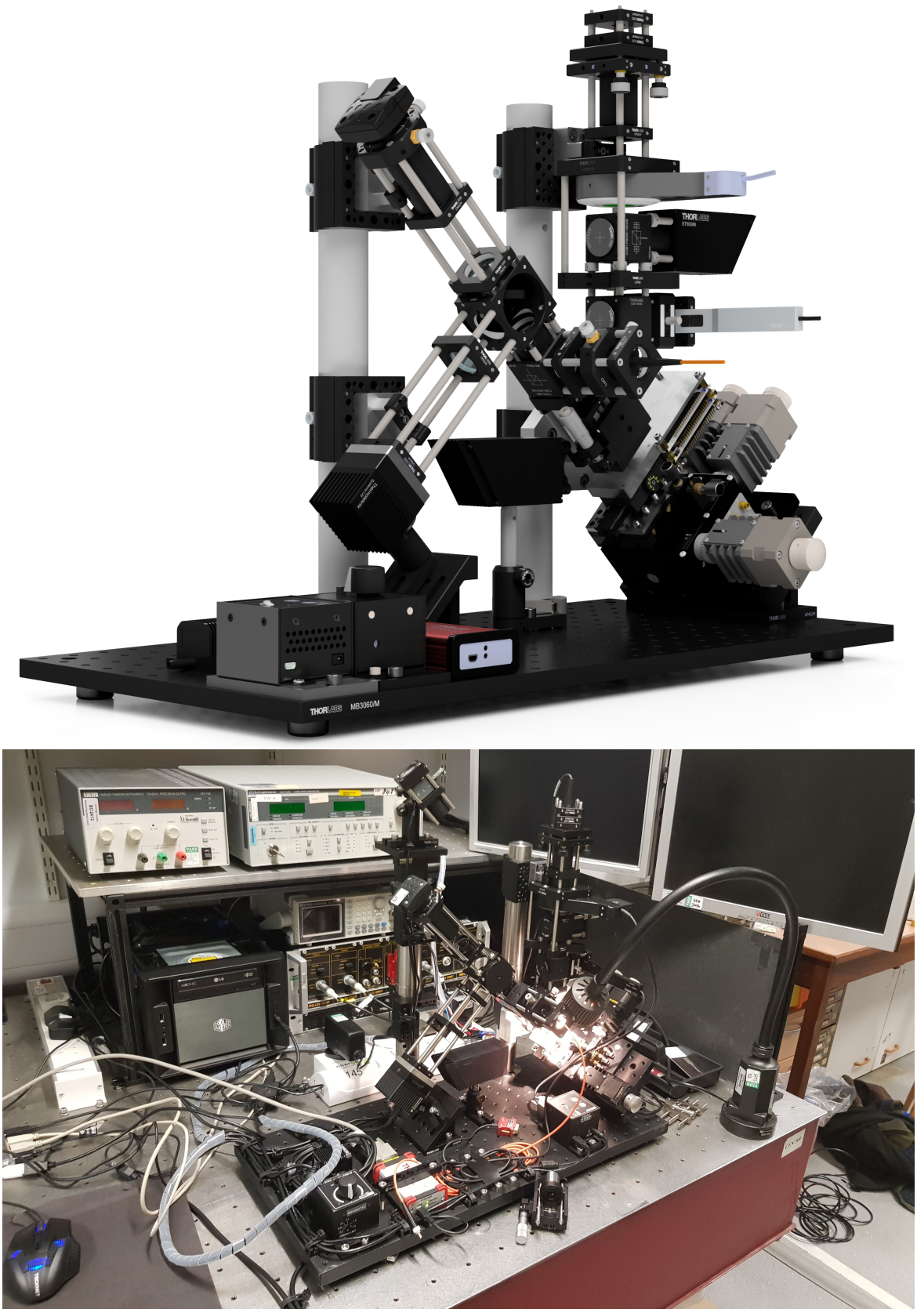


Fig. E.2 SLM power capacity experiment render (above) and photo (below)

The microscope system - blue in Figure E.1 - operates by passing a white light LED source through an objective and imaging the back reflected light. This allows for real time measurement of substrate degradation.

In order to ensure maximum flexibility, all components in the system are designed to be modular and interchangeable.

E.4 Automation, Control and Operation

A control suite for the system was developed in C# and C/C++ based on the HoloGen framework, Chapter 4. This is capable of automating the entire alignment, characterisation and metrology process with a minimum of initial user input.

E.4.1 Source Calibration

There are three automated calibration procedures for the source measuring power, stability and ellipticity.

Source Power Calibration

Calibration of the laser source power is straight forward provided the power sensor used is of known properties. The waveplate and polarising beam splitter are aligned with parallel fast axes and the response curve of source driving voltage to measured power is taken. Aligning the fast axis of the laser at 45° to the polarising beam splitter and repeating the measurement allows a second response curve to be measured. The combined response of the laser is equal to the sum of these measurements and allows us to determine laser power without removing the polarising beam splitter or half waveplate.

Source Stability Calibration

The stability of the laser source can be determined simply by holding the source at constant driving voltage and recording the change in measured power over a period of time, in this

case taken as a period of 8 hours. Taking sufficient measurements allows a least squares fit to a gaussian distribution in order to calculate the FWHM stability. In the application discussed below, stability was sufficient to ignore it from LIDT calculations.

As before, systems incorporating the waveplate and polarising beam splitter require two measurements at 45° in order to fully understand stability behaviour.

Source Ellipticity Calibration

When the waveplate is initially mounted at a non-zero angle θ_0 and the source is mounted with unknown orientation the waveplate is rotated through 360 degrees and the incident powers recorded. Linear regression then allows for determination of source properties from which the ellipticity can be determined. This is discussed in more detail in Appendix F.

E.4.2 Computer Vision

The initial focus of the microscope sub-system is set by the user. A basic software autofocus implementation is used with an integrated Zemax model of the objective lens system used to inform z-axis adjustments on the alignment stage. The power of the illumination LED is controlled to ensure good image white-balance and contrast and reduce post processing.

To automate the damage observation process, a control image is taken before the start of each test. After each test the recorded image I_i for measurement i is compared to the control image I_0 using a normalised mean squared error E_{MSE} where

$$E_{MSE}(I_0, I_i) = \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} [k|I_0(x,y)| - |I_i(x,y)|]^2 \quad \textit{where} \quad k = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \frac{|I_i(x,y)|^2}{|I_0(x,y)|^2} \tag{E.1}$$

and N_x and N_y are the respective x and y resolutions. A suitable cutoff value for E_{MSE} can then be taken. All captured images are preserved to allow for manual confirmation if required.

The system parameters are set by the user in a JSON format. These define the volume of operation for the stage as well as testing area on the component and initial power values for testing.

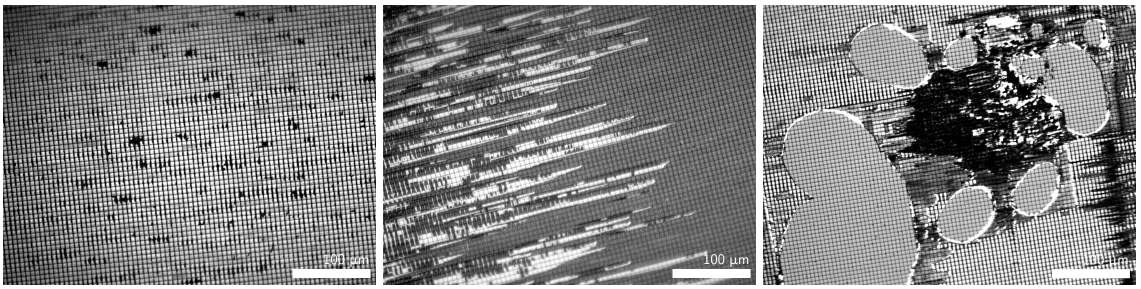


Fig. E.3 Optical microscope images of substrate damage using an Evolution MP 5.0 camera and Olympus BX60 microscope

E.5 Validation

The system was validated using NENIR30A ND filters from ThorLabs as they are low cost and have well defined damage thresholds under CW illumination. ThorLabs specify the NENIR30A as having a LIDT of 25 $\mathrm{W/cm^2}$ ⌀62µm at 1064nm while this system measured a LIDT of 29.6 $\mathrm{W/cm^2}$ ⌀70µm at 1090nm.

E.6 Demonstration

The experimental rig discussed so far was designed as part of ongoing research into SLM power handling capabilities and is demonstrated using a number of HDP-1280-2 'BlueJay' ferroelectric displays. These have a resolution of 1280×1280 pixels and a package size of 11mm by 25mm.

As SLMs are multi-level devices, the definition of 'damage' is taken to include any visible change in the device rather than simply visible change in the substrate.

As interest was in Near Infrared (NIR) behaviour, a $200W$ 1090 ± 5nm fibre laser source from was used. This is delivered to the system through a multi-mode fibre and the specifications are given in Table E.2.

Table E.2 Key SPI SP-200C specification details[312]

Parameter	Specification
Rated CW Output Power	200W
Polarisation	random
Wavelength	1090 ± 5nm
Power Stability over 8 Hours	$< 2\%$
Bandwidth	< 4nm
Beam Diameter ($1/e^2$)	5.0 ± 0.7mm
Full Angle Divergence	< 0.45mrad
M^2	1.1
Circularity	$> 90\%$
Eccentricity	$< \pm 1.2$mm
Concentricity	$< \pm 2.00$mrad

E.7 Results and Discussion

The automated system ran ≈ 350 tests for a number of spot sizes. The operator time for testing was under 35 minutes with a combined automated runtime of 6 hours. It is estimated that an entire operator day would be required in an equivalent manual system.

As can be expected, the measured maximum power was higher for smaller gaussian spot sizes with LIDTs of 9.2 $\mathrm{W/cm^2}$ ⌀27µm, 5.5 $\mathrm{W/cm^2}$ ⌀150µm and 3.2 $\mathrm{W/cm^2}$ ⌀3.1mm being measured at 1090 ± 5nm. This is presumed to be due to bulk heating.

A number of failure paradigms were observed with some extremal cases shown in Figure E.3. Figure E.3 (left) shows liquid crystal breakdown under prolonged exposure. Figure E.3 (centre) shows delamination of the liquid crystal from the glass without substrate damage and Figure E.3 (right) shows direct substrate damage. The captured microscope images were manually inspected with only one image being classified differently by the human operator and computer vision system.

While not unexpected, there was no observed difference in the LIDT against polarisation parallel or perpendicular to the SLM major axis.

E.8 Conclusion

This work has presented a fully automated system for Laser Induced Damage Threshold testing of substrates using only commercial off-the-shelf components. The setup requires $< 10\%$ of the operator time required for the equivalent manual system and reduces the manual error sources.

The system was demonstrated by testing a Liquid Crystal on Silicon (LCoS) device. LIDTs of 9.2 $\mathrm{W/cm^2}$ ⌀27µm, 5.5 $\mathrm{W/cm^2}$ ⌀150µm and 3.2 $\mathrm{W/cm^2}$ ⌀3.1mm were found for the active device face with an excitation wavelength of 1090 ± 5nm.

Appendix F

Ellipticity Measurement

Appendix F

Ellipticity Measurement

This appendix gives a description of the system used for calibrating the ellipticity of the various lasers used in Appendix E.

Section 2.1.2 introduced Jones Matrices. An arbitrary elliptical polarisation state can be represented as

$$E \begin{bmatrix} A \\ Be^{i\delta} \end{bmatrix} \tag{F.1}$$

where $\sqrt{A^2+B^2} = 1$ and A, B, E and δ are scalar constants.

Figure 2.4 introduced the Polarisation Ellipse and the Poincaré Sphere. The relationship between δ in (F.1) and ψ in Figure 2.4 is given by

$$\psi = \frac{1}{2}\tan^{-1}\left(\frac{2AB\cos\delta}{A^2-B^2}\right) \tag{F.2}$$

The minimum and maximum values are given by [313]

$$E_{\psi} = E\sqrt{A^2\cos^2\psi + B^2\sin^2\psi + AB\cos\delta\sin 2\psi} \tag{F.3}$$

$$E_{\psi\pm\frac{\pi}{2}} = E\sqrt{A^2\sin^2\psi + B^2\cos^2\psi - AB\cos\delta\sin 2\psi} \tag{F.4}$$

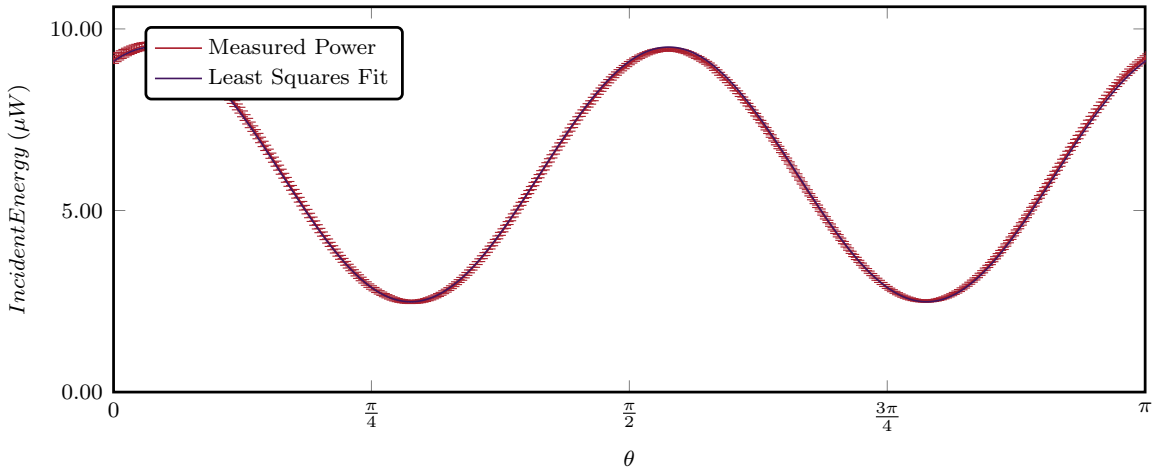


Fig. F.1 Power incident on the power meter vs waveplate angle

For the system in hand with inputs from Table 2.1

$$\begin{aligned} E_{pm} &= C_{bs}C_{nd}\left|\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} e^{\frac{-i\pi}{2}} \begin{bmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{bmatrix} E \begin{bmatrix}A\\ Be^{i\delta}\end{bmatrix}\right| \\ &= \sqrt{C_{bs}}\sqrt{C_{nd}}Ee^{-\frac{i\pi}{2}}\left|A\cos 2\theta + Be^{i\delta}\sin 2\theta\right| \\ &= \sqrt{C_{bs}}\sqrt{C_{nd}}Ee^{-\frac{i\pi}{2}}\sqrt{\frac{4AB\cos\delta\cos 2\theta\sin 2\theta + (A^2+B^2) + (A^2-B^2)\cos 4\theta}{2}} \end{aligned} \quad \text{(F.5)}$$

where constants C_{bs} and C_{nd} represent loss of power due to beam splitter and neutral density (ND) filter respectively and E_{pm} is the measured intensity at the power meter.

In practice, the waveplate is initially mounted at a non-zero angle θ_0 and the diode is mounted with unknown orientation. The waveplate is rotated through 360 degrees and the incident powers recorded. An example is shown in Figure F.1.

Measured intensity I_{pm} is equal to $\frac{1}{2}nc\varepsilon_0 E_{pm}^2$[314]. Collecting terms in (F.5) gives [1]

$$I_{pm} = \frac{nc\varepsilon_0}{C_{bs}C_{nd}} \frac{E_{pm}^2}{E^2}$$

$$I_{pm} = CE_{pm}^2, \quad \text{where} \quad C = \frac{nc\varepsilon_0}{C_{bs}C_{nd}E^2}$$

$$= \underbrace{C\frac{\left(A^2+B^2\right)}{2}}_{\text{Constant Term}} + \underbrace{C\sqrt{\frac{\left(A^2-B^2\right)^2 + A^2B^2\cos\delta^2}{2}}}_{\text{Amplitude Term}} \underbrace{\sin\left(4\theta_0 + 4\theta + \tan^{-1}\left(\frac{\left(A^2-B^2\right)}{2AB\cos\delta}\right)\right)}_{\text{Frequency Term}} \tag{F.6}$$

A least squares regression to a curve of form $a_1 \sin x + a_2 \cos x + a_3$ allows an accurate model of this to be retrieved from the measured powers. Elementary trigonometric identities can then be used to find accurate values of any three of A, B, δ and θ_0 (or E_ψ, $E_\psi \pm \frac{\pi}{2}$, δ and θ_0) provided the other is known with no manual intervention in the system. In the case of the 980 nm diode laser shown in Figure F.1, the half waveplate had $\theta_0 = 82^\circ$. This leads to

$$\delta = 1.16, \qquad A = 0.86 \tag{F.7}$$

$$\psi = 17.76^\circ, \qquad B = 0.54 \tag{F.8}$$

$$E_{max} = \max\left(E_\psi, E_{\psi\pm\frac{\pi}{2}}\right) \qquad = 3.08 \tag{F.9}$$

$$E_{min} = \min\left(E_\psi, E_{\psi\pm\frac{\pi}{2}}\right) \qquad = 1.57 \tag{F.10}$$

$$e = \frac{\max\left(E_\psi, E_{\psi\pm\frac{\pi}{2}}\right)}{\min\left(E_\psi, E_{\psi\pm\frac{\pi}{2}}\right)} \qquad = 1.95 \tag{F.11}$$

where e is the *ellipticity*.

[1] Using $a\sin x + b\cos x = c\sin(x+\phi)$ where $c = \sqrt{a^2+b^2}$ and $\phi = tan^{-1}\left(\frac{b}{a}\right)$

Appendix G

HoloGen Application Architecture

Appendix G

HoloGen Application Architecture

G.1 Introduction

Chapter 4 presented a brief summary of the application library architecture but deferred from further analysis. While the majority of HoloGen is simple contextually and requires little discussion, a few areas stand out as meriting explanation for future developers. Several key areas are discussed: The parameter and command tree in Section G.2; The native algorithm interface in Section G.3; The serialisation architecture in Section G.4 and the user interface construction in Section G.5. Section G.6 discusses the role of all the application libraries and their connections as well as any key classes not covered elsewhere in this appendix. Finally Section G.7 provides some areas of future work and the infrastructure put in place to handle them.

G.2 Parameter and Command Hierarchy

HoloGen uses a custom reflection based parameter and command system. This is in contrast to the XML parameter sheet systems widely in use. Instead of the parameter types and interactions being defined in parameter sheets which are parsed at runtime, the parameter system is coded into the C# directly. This significantly reduces the runtime overhead as well as improves the error checking available at compile time. The downside is an increased

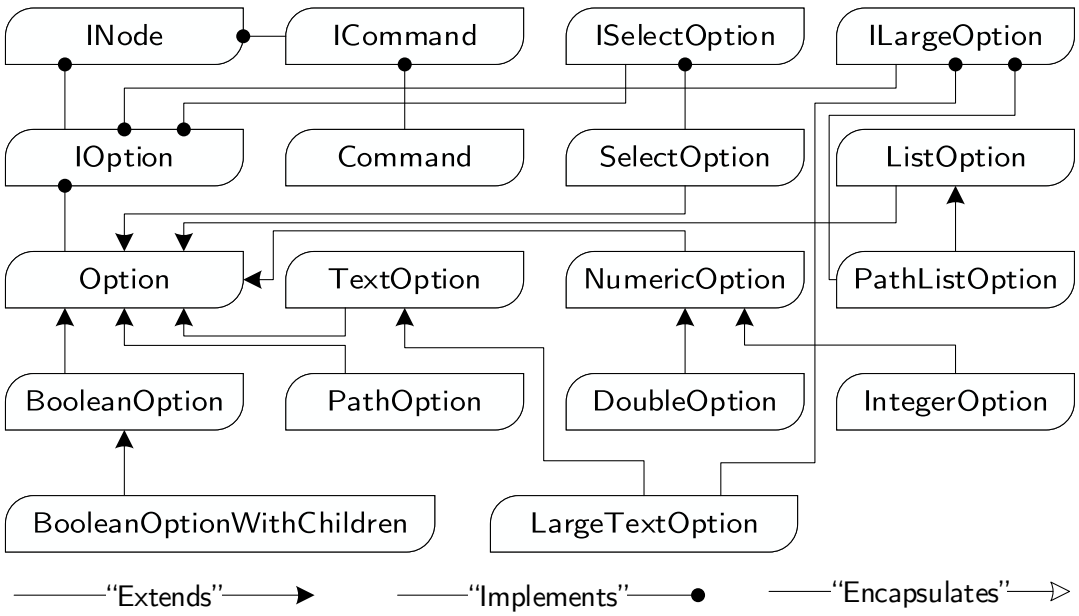


Fig. G.1 HoloGen parameter types inheritance hierarchy

architecture exposure of the parameter hierarchy. The decision to use the reflection based system was made to enable a side project of the author.

G.2.1 Class Inheritance Hierarchy

Two distinct areas of the *HoloGenHierarchy* library code stand out: the parameter/command types (e.g. numerical options, menu commands, etc) and the tree of elements containing them (e.g. pages, menus, etc).

The parameter types are defined using the following key classes shown in Figure G.1:

- `INode` - Base interface for all *leaf* nodes in the parameter hierarchy.
- `ICommand` - Base command interface extended by all parameter types. Extends `INode`.
- `IOption` - Base parameter interface extended by all parameter types. Extends `INode`.
- `Command` - Abstract base command class extended by all command types. Distinct from `ICommand` which it implements due to C# not handling generic template references.

- `Option` - Abstract base parameter class extended by all parameter types. Distinct from `IOption` which it implements due to C# not handling generic template references. A template system allows for generic manipulation of wrapped values without exposing the internals of the class to extending objects.
- `NumericOption` - Abstract base parameter class extended by all numeric parameter types. Extends `Option`.
- `IntegerOption` - Integral numeric parameter type. Extends `NumericOption`.
- `DoubleOption` - Floating point numeric parameter type. Extends `NumericOption`.
- `ILargeOption` - Interface that flags to the display that implementing options require extra space on the UI. Extends `IOption`.
- `TextOption` - Text based parameter type. Extends `Option`.
- `LargeTextOption` - Larger version of `TextOption`. Implements `ILargeOption`.
- `PathOption` - Alternative to `TextOption` that handles file paths. Implements `ILargeOption`.
- `ISelectOption` - Base parameter interface extended by all selection based parameter types. Extends `IOption`.
- `SelectOption` - Abstract base parameter class extended by all selection based parameter types. Distinct from `ISelectOption` which it implements due to C# not handling generic template references. A template system allows for generic manipulation of wrapped values without exposing the internals of the class to extending objects. Contains a `PossibilityCollection` of `Possibilities` that can be selected as the option value. Any options owned by the selected `Possibility` are injected into the owning `HierarchyFolder`.
- `ListOption` - Base class for parameters representing lists of values. Extends `Option`. Implements `ILargeOption`.
- `PathListOption` - Base class for parameters representing lists of files. Extends `ListOption`. Implements `ILargeOption`.
- `BooleanOption` - Boolean value based parameter type. Extends `Option`.
- `BooleanOptionWithChildren` - Boolean value based parameter type. Extends `BooleanOption`. When set to *true*, any child options are injected into the owning `HierarchyFolder` in a manner similar to `SelectOption`.

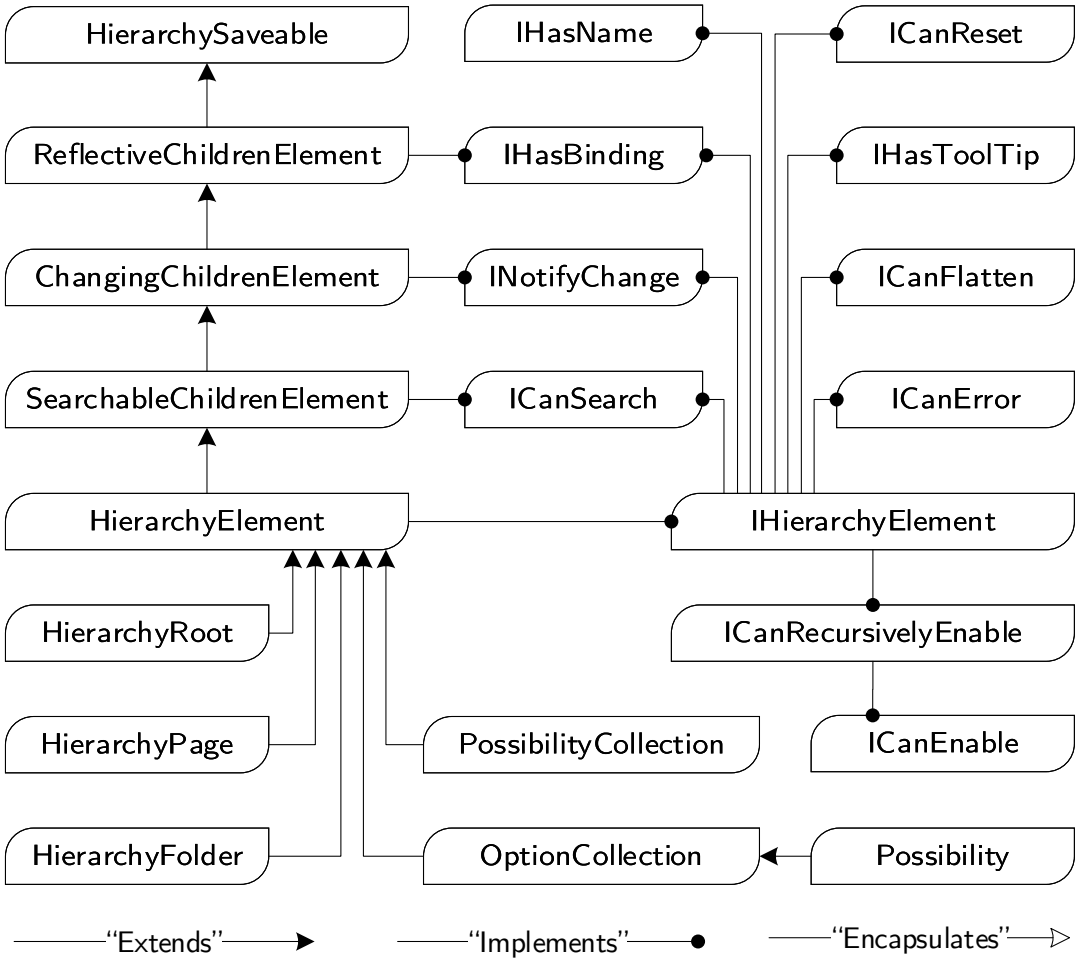


Fig. G.2 HoloGen paramater tree inheritance hierarchy

The tree of elements containing the parameters or commands are shown in Figure G.2:

- `ReflectiveChildrenElement` - Key class representing any *non-leaf* node in the parameter/command tree. When extended, this class uses the C# reflection system to find all public parameters of the same type as the specified template type and uses them to populate a list of children of that type. This allows extending classes to declare member elements without having to handle their manipulation or access.
- `ChangingChildrenElement` - Extends `ReflectiveChildrenElement` to handle a *non-leaf* node that has a changing set of children with appropriate notifications.

- `ChangingChildrenElement` - Extends `ChangingChildrenElement` to handle a *non-leaf* node that has a set of children which can be searched.
- `IHierarchyElement` - Abstract interface for all *non-leaf* nodes in the parameter tree.
- `HierarchyElement` - Abstract base class for all *non-leaf* nodes in the parameter/command tree. Distinct from `IHierarchyElement` which it implements due to C# not handling generic template references.
- `HierarchyRoot` - Implementation of `HierarchyElement` that represents the root node of a parameter/command tree. Equivalent to the *tab* or *menu pop-out* level within HoloGen.
- `HierarchyPage` - Implementation of `HierarchyElement` that represents a node of a parameter/command tree. Equivalent to the *page* or *menu* level within HoloGen.
- `HierarchyFolder` - Implementation of `HierarchyElement` that represents a node of a parameter/command tree. Equivalent to the *folder* or *sub-menu* level within HoloGen.
- `OptionCollection` - Implementation of `HierarchyElement` that represents a set of `Options` within a specialisation `Option` such as `BooleanOptionWithChildren` or `SelectOption`.
- `PossibilityCollection` - Implementation of `HierarchyElement` that represents a set of `Possibilities` within a `SelectOption`.
- `Possibility` - Represents a possibility state for a `SelectOption`.
- `HierarchyVersion` - Represents a version number for a parameter/command hierarchy.
- `HierarchySaveable` - Represents any class that can be saved using the JSON serialisation.

In addition a number of function interfaces are used to mark exhibited behaviours at different levels of the hierarchy.

- `ICanEnable` - Implemented by any class/interface that exhibits enable/disable behaviours.
- `ICanError` - Implemented by any class/interface that can be in an error state.
- `ICanImportFromString` - Implemented by any class/interface that allows for deserialisation from a string object.

- `ICanExportToString` - Implemented by any class/interface that allows for serialisation to a string object.
- `ICanFlatten` - Implemented by any `HierarchyElement` that can flatten its internal tree structure.
- `IHasName` - Implemented by any class/interface that has a name property.
- `IHasToolTip` - Implemented by any class/interface that has a tool tip property.
- `IHasWatermark` - Implemented by any class/interface that has a watermark property.
- `IHasBindingPath` - Implemented by any `HierarchyElement`, `Command` or `Option` that the interface can be bound to given a link to the base of the parameter or command tree.
- `ICanSearch` - Implemented by any class/interface that can be searched.
- `ICanReset` - Implemented by any class/interface that can be reset.
- `ICanRecursivelyEnable` - Implemented by any class/interface that can set its own enabled state and that of its children. Extends `ICanEnable` .
- `INotifyChanged` - Implemented by any `Option` that notifies when the contained value changes.

G.2.2 Application

HoloGen defines a number of parameter and command hierarchies including those in the *HoloGenOptions*, *HoloGenImageOptions*, *HoloGenProcessOptions* and *HoloGenBatchOptions* libraries. These follow a standard structure with the following classes.

- A root element extending `HierarchyRoot` that provides a name and tool tip as well as containing public properties for all of the `HierarchyPage` objects within. Example: `OptionsRoot` in the *HoloGenOptions* library.
- Multiple elements extending `HierarchyPage` that provides a name, tool tip and icon as well as containing public properties for all of the `HierarchyFolder` objects within. Example: `ProjectorPage` in the *HoloGenOptions* library.

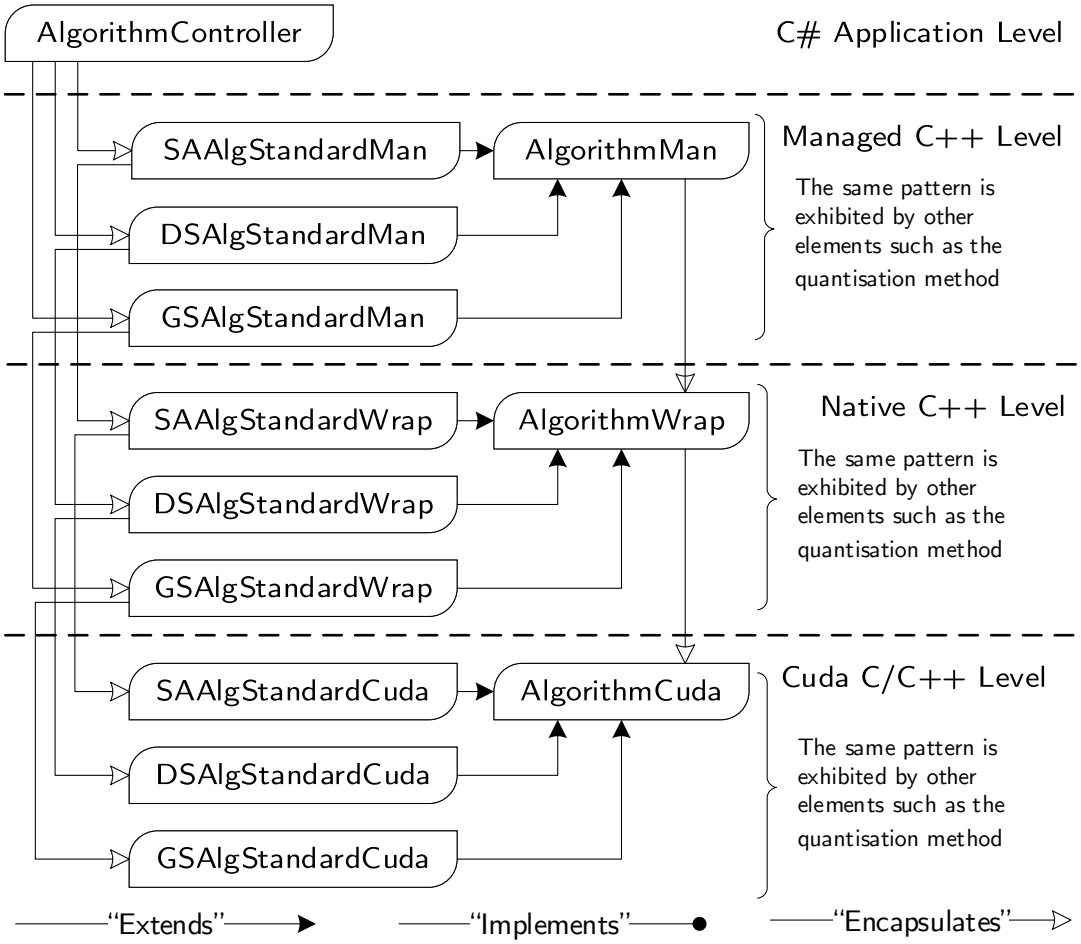


Fig. G.3 HoloGen algorithm interface

- Multiple elements extending `HierarchyFolder` that provides a name and tool tip as well as containing public properties for all of the `Option` objects within. Example: `HologramFolder` in the *HoloGenOptions* library.
- Multiple elements extending `Option` or its subclasses that provide the name, tool tip, defaults and limits for the option. Example: `SLMResolutionX` in the *HoloGenOptions* library.
  - Classes extending `SelectOption` contain a link to an extension of a `PossibilityCollection` object. Example: `SLMTypeOption` in the *HoloGenOptions* library.x
  - Classes extending `Possibility` or `BooleanOptionWithChildren` are also able to contain public properties that will only be editable when selected. Example: `MultiAmpSLM` in the *HoloGenOptions* library.

- Elements extending `PossibilityCollection` with public properties for each of the allowable `Possibilities` of the `SelectOption`. Example: `SLMPossibilities` in the *HoloGenOptions* library.
- Elements extending `Possibility` with public properties for each of the allowable `Possibilities` of the `SelectOption`. Example: `SLMPossibility` in the *HoloGenOptions* library.

While the reflection based architecture requires an initial investment of time and effort to comes to grips with, it has proved highly time efficient in practice during development on HoloGen. The system presents an alternative to the XML systems commonly used and should be considered for wider use.

G.3 Algorithm Interface

A four step process is required in order to pass data from the C# HoloGen application and user interface to the Cuda C/C++ underlying it. Figure G.3 shows this for a couple of example cases.

The top level is the C# application level where an `AlgorithmController` unpacks the `Option` parameter hierarchy and passes it to the Managed C++ level below. All Managed C++ wrapper classes have the suffix "*Man". Managed C++libraries are able to use both native data types as well as the *.NET* data types used by C#. After the data is copied between the two types, the native C++ layer is called. This level can be exposed in a dynamic or static library but is not able to link to many native Cuda headers such as CUFFT and Thrust. All native C++ wrapper classes have the suffix "*Wrap". These classes can, in turn, pass the data onto Cuda C/C++ compiled classes which are able to communicate with the graphics card. All Cuda C/C++ classes have the suffix "*Cuda".

For example, when passing a target image to a GS algorithm, the following steps occur. `AlgorithmController` calls `SetTargetImage()` on a `GSAlgStandardMan` object it owns. This uses its base `Convert)(GSAlgStandardMan)` functions inherited from `AlgorithmMan` to transfer the data to native types. This is then passed to the `GSAlgStandardWrap` and `GSAlgStandardCuda` objects in turn. Once in the Cuda level, the parent functions in `AlgorithmCuda` handle transferring the data to the graphics card and maintaining handles on its location. The actual algorithm implementations can be found in the `RunIterations()`

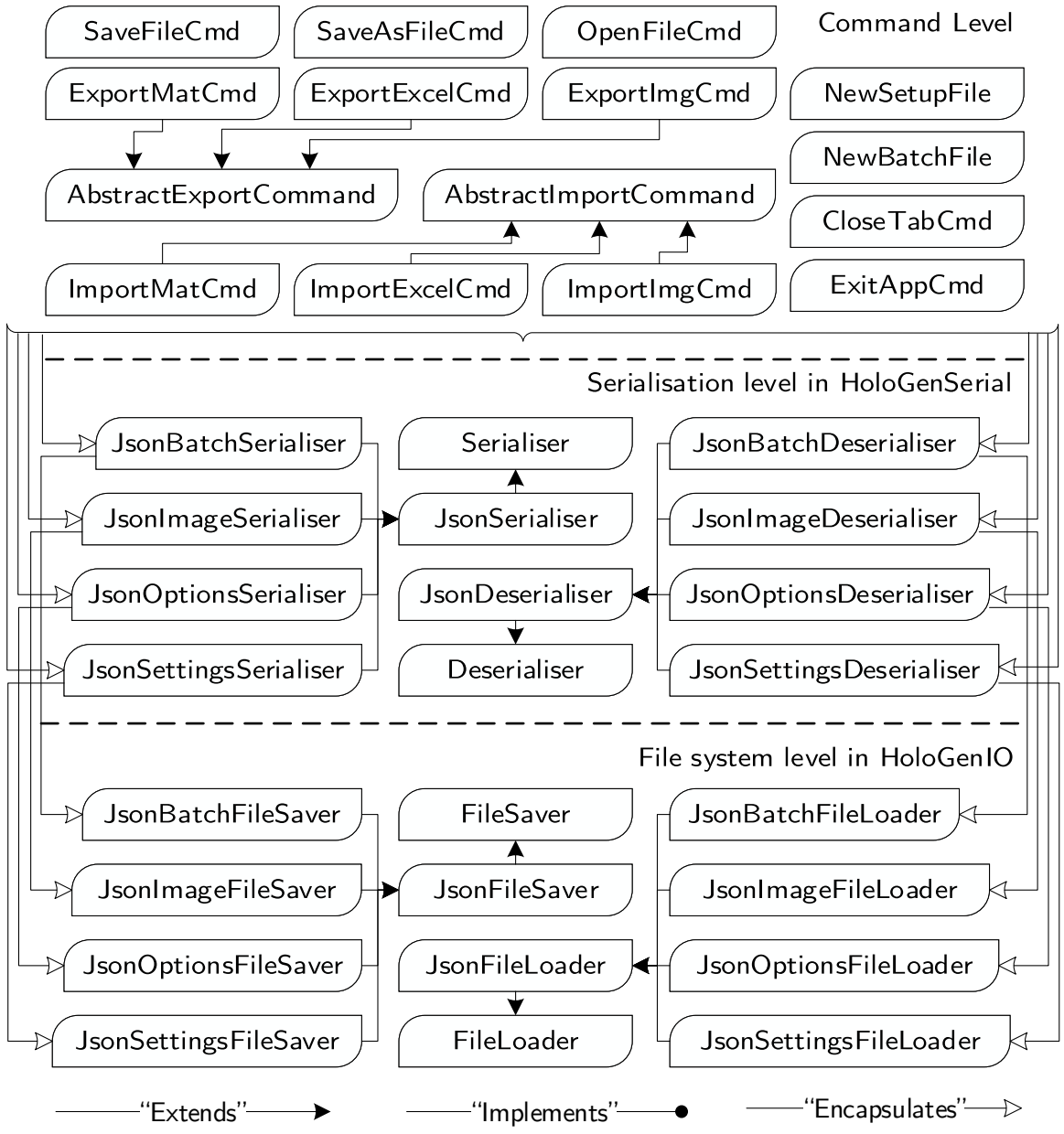


Fig. G.4 HoloGen serialisation architecture

functions of their respective "*Cuda" classes. These are comparatively straightforward and can be compared with those discussed in Chapter 2 and shown in Algorithms 2.1, 2.2, 2.3, 2.4 and 2.5. Their relative performance is discussed in Appendix B.

G.4 Serialisation

Figure G.4 shows the serialisation architecture for HoloGen. The command level in libraries such as *HoloGenUIMenu* interface to the JSON serialisation and deserialisation classes in the *HoloGenSerial* library. These then write to and read from the file system using the classes found in the *HoloGenIO* library.

G.5 User Interface

HoloGen follows a standard *View-Model-ViewModel* structure for its user interface. Each UI element or *View* is defined graphically in a *.xaml file, e.g. `SetupTabView.xaml`, with a C# companion file, e.g. `SetupTabView.xaml.cs`. The contained UI elements bind to the data contained and manipulated by the *ViewModel*, e.g. `SetupTabViewModel.cs`, which in turn holds handles to the internal data, e.g. `OptionsRoot.cs`.

This paradigm is widely used in the software world and it is beyond the scope of this document to expand on it further.

G.6 Library Descriptions

Figure G.5 shows the layout of libraries within HoloGen with reference to the three application levels shown in Figure 4.7. This section discusses the libraries used and any key classes not discussed in earlier sections.

G.6.1 User Interface Libraries

All user interface libraries are prefixed with "HoloGenUI" and are written in C# using WPF.

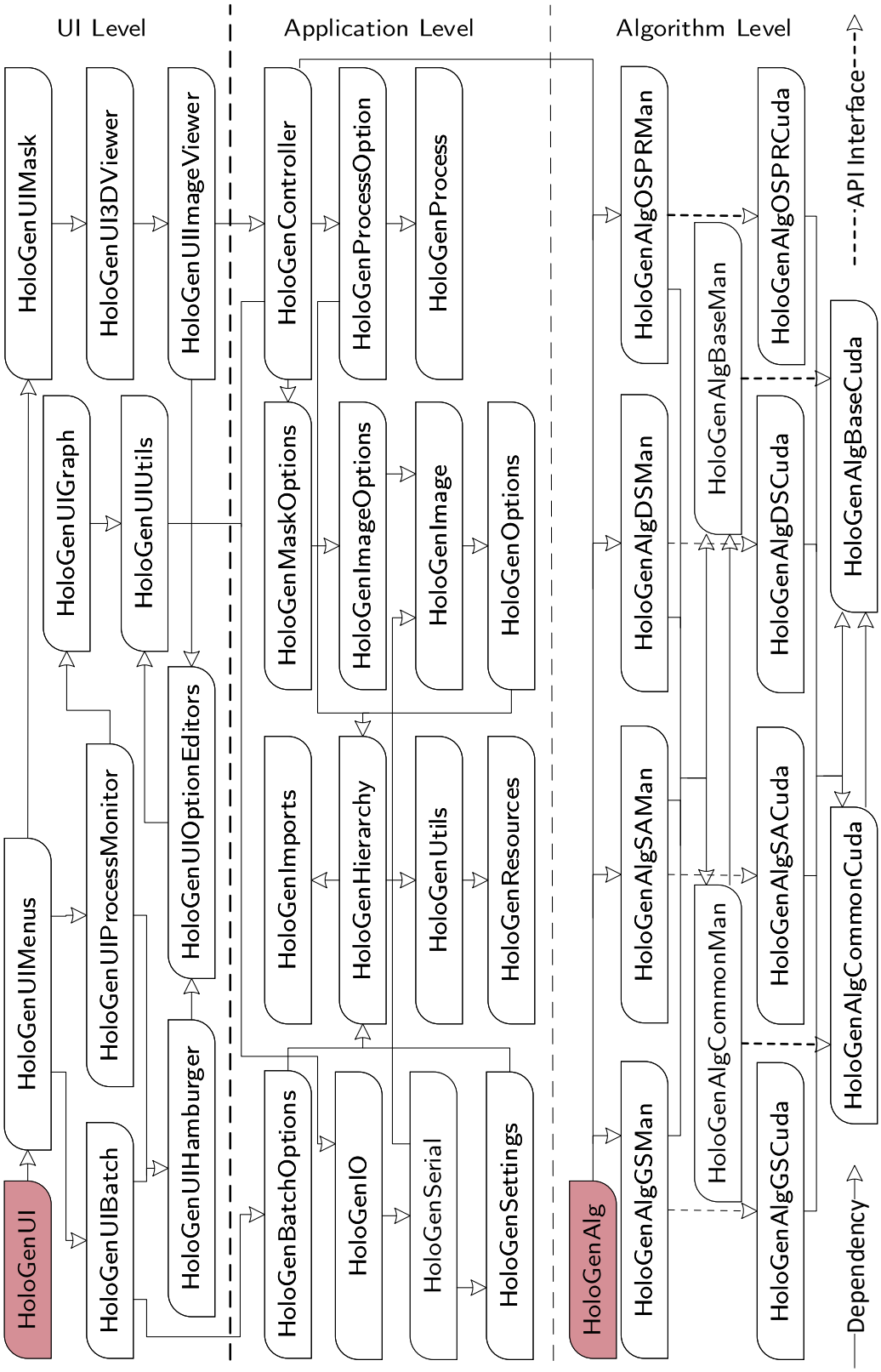


Fig. G.5 HoloGen package layout.

- *HoloGenUI* - The main entry point for the user application user interface. Provides top-level implementation `TabHandlers` for in *HoloGenUIUtils*. Key classes include:
  - `MainWindowViewModel` - ViewModel construct underlying the main application window. Keeps an observable collection of `AbstractTab` objects and is responsible for matching them with appropriate display classes. Implements `ITabHandler` for the `TabHandlerFramework`. `MainWindowViewModel` is responsible for the application wide user settings including notifications and file IO. Drag and drop behaviours are also overseen by `MainWindowViewModel` using the *Dragablz* library.
  - `AbstractTab` - Tab data interface that all other tabs can inherit. This allows for templating using `TabDataTemplateSelector`.
  - `SetupTabViewModel` - Concrete implementation of `AbstractTab` for setting up a hologram generation process. `SetupTabViewModel`'s main function serves to wrap a `HamburgerTabViewModel` object from *HoloGenUIHamburger*.
  - `BatchTabViewModel` - Concrete implementation of `AbstractTab` for batch processing data. `BatchTabViewModel`'s main function serves to wrap a `BatchViewModel` object from *HoloGenUIBatch*.
  - `BrowserTabViewModel` - Concrete implementation of `AbstractTab` providing a built in browser. `BrowserTabViewModel`'s main function serves to wrap the imported *CEFSharp* functionality.
  - `ProcessTabViewModel` - Concrete implementation of `AbstractTab` for visualising running process parameters and reporting messages from the underlying algorithms. `ProcessTabViewModel`'s main function serves to wrap a `ProcessViewModel` object from *HoloGenUIProcessMonitor*.
  - `HologramTabViewModel` - Concrete implementation of `AbstractTab` for visualising complex valued images in 2d and 3D with optional masking data.. `HologramTabViewModel`'s main function serves to wrap a `MaskViewModel` object from *HoloGenUIMask*.

  The display code is written in C# and WPF and lies on top of the ViewModels. There is a one to one correspondence between ViewModel classes and XAML parameter sheets on top.
- *HoloGenUIMenus* - Defines `Options` and `Commands` for the main application flyout menu. Built on-top of the *HoloGenHierarchy* library discussed in Section G.2.

- *HoloGenUIHamburger* - Defines a series of View/ViewModel pairs that serve to unpack and display *HoloGenHierarchy* parameter hierarchies in a Microsoft Metro "Hamburger Menu" style. Individual editors are defined in *HoloGenUIOptionEditors*.
- *HoloGenUIOptionEditors* - Defines a series of View/ViewModel pairs that serve to display the different *Option* types discussed in Section G.2. Primarily used as the *leaf nodes* for *HoloGenUIHamburger*.
- *HoloGenUIProcessMonitor* - Defines display classes for the algorithm monitoring tab. The charting is built on top of *LiveCharts* using *HoloGenUIGraph*.
- *HoloGenUIGraph* - Defines display classes for realtime charting. *HoloGenUIGraph* is built on top of *LiveCharts*.
- *HoloGenUIResources* - Defines translation strings for the application localisation.
- *HoloGenUIBatch* - Defines display classes for the batch processing tab. Defines a number of *AbstractColumnFactory* implementations for displaying different `Option` types. Key classes include:
  - `AbstractColumnFactory` - Interface for a series of factory objects - `BooleanColumnFactory`, `DefaultColumnFactory`, `DoubleColumnFactory`, `IntegerColumnFactory`, `PathColumnFactory`, `SelectColumnFactory`, and `TextColumnFactory` - used for generating WPF `DataGridColumns`.
  - `TemplateGenerator` - Helper class for `AbstractColumnFactory` implementations that uses a given delegate to create new instances for similar `Option` types.
- *HoloGenUIImageViewer* - Defines classes for viewing complex valued images in 2D.
- *HoloGenUIMask* - Extends the functionality from *HoloGenUIImageViewer* showing masked regions on a 2D complex valued image.
- *HoloGenUI3DViewer* - Extends the functionality from *HoloGenUIMask* showing complex valued images in 3D.
- *HoloGenUIUtils* - Defines low level utility classes for the HoloGen UI as well as abstract interfaces for services offered by higher level libraries. Key classes include:
  - `ITabHandler` and `TabHandlerFramework` - Define a service framework. The `MainWindowViewModel` extends `ITabHandler` and can register itself with `TabHandlerFramewo` Lower level class libraries can then call the `TabHandlerFramework` while remaining ignorant of the implementation.

  - `CreatorsThesis` - Display theme for the application built on top of *MahApps*.
  - `ICanExport`, `ICanExportBitmap`, `ICanExportExcel`, `ICanExportMat` and `ICanSave` - Interfaces that `ITabHandler` implementations can extend to declare their IO requirements.

G.6.2 Application Libraries

All application libraries are prefixed with "HoloGen".

- *HoloGenController* - Defines classes that unwrap the `Options` hierarchy defined in *HoloGenOptions* for hologram generation and communicate it with the algorithms wrapped by *HoloGenAlgBaseMan*, *HoloGenAlgGSMan*, *HoloGenAlgSAMan*, *HoloGenDSMan* and *HoloGenOSPRMan*. Key class is `AlgorithmController`.
- *HoloGenProcess* - Defines data structures for defining a hologram generation process.
- *HoloGenImage* - Defines classes related to complex valued images. Key classes include:
  - `ComplexImage` - Object that holds an image in `Complex` format as well as generation metadata and pre-cached values. Defines specialised JSON interface commands in order to preserve disk space.
  - `ImageCache` - Defines a cache for different `Bitmap` views on a `ComplexImage` object. Once a particular visualisation bitmap is generated for a particular image, the result is cached to reduce future load times. Uses a `GenericCache` from *HoloGenUtils* as the underlying implementation with the `TransformType`, `ImageViewType`, `ColorScheme` and `ImageScaleType` enums as the four access keys.
- *HoloGenHierarchy* - Defines the `Options` and `Command` hierarchy discussed in Section G.2.
- *HoloGenImports* - Classes or constructs that have been imported from other applications in a source code format.
- *HoloGenResources* - Defines translation strings for the application localisation.
- *HoloGenOptions* - Defines an `Options` hierarchy for the hologram generation algorithms. Built on-top of the *HoloGenHierarchy* library discussed in Section G.2.

- *HoloGenProcessOptions* - Defines an `Options` hierarchy for manipulating the running process display in *HoloGenUIProcessMonitor*. Built on-top of the *HoloGenHierarchy* library discussed in Section G.2.
- *HoloGenBatchOptions* - Defines an `Options` hierarchy for the batch processing hologram generation algorithms in *HoloGenUIBatch*. Built on-top of the *HoloGenHierarchy* library discussed in Section G.2.
- *HoloGenSettings* - Defines an `Options` hierarchy for the application settings in *HoloGenUI* and *HoloGenUIMenus*. Built on-top of the *HoloGenHierarchy* library discussed in Section G.2.
- *HoloGenImageOptions* - Defines an `Options` hierarchy for visualising complex valued images in *HoloGenUIImageViewer*. Built on-top of the *HoloGenHierarchy* library discussed in Section G.2.
- *HoloGenMaskOptions* - Defines an `Options` hierarchy for visualising complex valued images with masking data in *HoloGenUIMask*. Built on-top of the *HoloGenHierarchy* library discussed in Section G.2.
- *HoloGenSerial* - Handles the serialisation of `HierarchySaveable` data objects. Discussed further in Section G.5.
- *HoloGenIO* - Handles the file input an doutput of serialised of `HierarchySaveable` data objects. Discussed further in Section G.5.
- *HoloGenUtils* - Utility classes for HoloGen at the application level. Key classes include:
  - `ComplexImage` - Object that holds an image in `Complex` format as well as generation metadata and pre-cached values. Defines specialised JSON interface commands in order to preserve disk space.
  - `GenericCache` - Defines a cache for computationally expensive results using up to four different keys. Lambda functions are used for the computation to increase reusability. Provides the base implementation for `ImageCache`.
- *HoloGen* - Command line interface for the HoloGen application.

G.6.3 Algorithm Libraries

All algorithm libraries are prefixed with "HoloGenAlg" and recieve the "Man" or "Cuda" suffix depending on whether they are written in Managed C++ or Cuda C/C++.

- *HoloGenAlg* - Command line interface for the HoloGen algorithms. Used for batch processing test data in a client-server configuration.
- *HoloGenAlgBaseMan* - Managed C++ wrapper for *HoloGenAlgBaseCuda*.
- *HoloGenAlgCommonMan* - Managed C++ wrapper for *HoloGenAlgCommonCuda*.
- *HoloGenAlgGSMan* - Managed C++ wrapper for *HoloGenAlgGSCuda*.
- *HoloGenAlgSAMan* - Managed C++ wrapper for *HoloGenAlgSACuda*.
- *HoloGenAlgDSMan* - Managed C++ wrapper for *HoloGenAlgDSCuda*.
- *HoloGenAlgOSPRMan* - Managed C++ wrapper for *HoloGenAlgOSPRCuda*.
- *HoloGenAlgBaseCuda* - C++ and Cuda C base level algorithm definitions. Discussed further in Section G.3. Key classes include:
  - `AlgorithmCuda` - Base level algorithm definition. Stores target, illumination and starting images as well as common algorithm parameters.
  - `FFTHandlerCuda` - Wraps the *CUFFT* FFT library with *Thrust* friendly functions.
  - `FFTUpdaterCuda` - Implements (2.72) for updating the replay field after the change of a single refraction field pixel.
  - `Normaliser` - Provides a fast vector normalisation feature.
  - `Randomiser` - Allows for randomisation of different combinations of phase and amplitude.
  - `QuantiserCuda` - Base level quantisation algorithm.
  - `StridedChunkRange` - Customised *Thrust* iterator that can iterate through square regions of a matrix.
- *HoloGenAlgCommonCuda* - C++ and Cuda C common utility functions.
- *HoloGenAlgGSCuda* - C++ and Cuda C implementation of the Gerchberg-Saxton algorithm and other iterative Fourier transform approaches. See Section 2.4 for a discussion of this algorithm's operation.

- *HoloGenAlgSACuda* - C++ and Cuda C implementation of the simulated annealing algorithm. See Section 2.4 for a discussion of this algorithm's operation.
- *HoloGenAlgDSCuda* - C++ and Cuda C implementation of the direct binary search algorithm. See Section 2.4 for a discussion of this algorithm's operation.
- *HoloGenAlgOSPRCuda* - C++ and Cuda C implementation of the One-Step Phase-Retrieval algorithm. See Section 2.4 for a discussion of this algorithm's operation.

G.6.4 Imported Libraries

HoloGen also uses a number of imported libraries. A brief description of their function is presented here and the licensing information is presented in Chapter 4.

- *MahApps* - Custom controls for WPF apps as well as a material design skin based on Microsoft's Metro UI.
- *LiveCharts* - Customisable and bindable real-time charting library.
- *Dragablz* - Draggable tabs for WPF.
- *ControlzEx* - Custom controls for WPF apps.
- *HelixToolkit* - 3D viewer for WPF apps.
- *Xamarin.Forms* - Mobile/tablet compatibility.
- *MaterialDesign* - Material Design compatible skin for WPF apps.
- *MaterialSkin* - Alternative Material Design compatible skin for WPF apps.
- *GongSolutions.WPF.DragDrop* - WPF drag and drop capability.
- *FastMember.Signed* - Fast reflection for *.NET*.
- *DocumentFormat.OpenXML* - Microsoft Office file format interoperability.
- *ExcelNumberFormat* - Advanced Excel number formatting.
- *SharpDX* - WPF DirectX compatibility.
- *ClosedXML* - Microsoft Excel integration.

- *CefSharp* - WPF compatibly wrapper for the Chromium browser.
- *Cuda* - Programmable interface to NVidia graphics cards.
- *PdfiumViewer* - PDF file viewer.
- *Xceed* - DataGrid controls for WPF apps.
- *NHotKey* - Global hot keys for WPF apps.
- *Newtonsoft.Json* - JSON serialisation for .NET languages.
- *MathNet* - Mathematics package for .NET languages.
- *AForge* - Mathematics package for .NET languages.
- *Accord* - Mathematics package for .NET languages.
- *NUnit* - unit test framework for .NET languages.

G.7 Future Work

While built primarily as a research tool, HoloGen was designed with future expansion and commercialisation in mind. A number of features were planned but never executed. While these are discussed in more detail in the project GitLab pages, they are covered here for completeness.

- **More algorithm variants** - HoloGen was deliberately designed to allow for the easy addition and recomposition of algorithms and its library of available techniques should be expanded.
- **More performance metrics** - Currently HoloGen only exposes mean squared error (MSE) and efficiency metrics for algorithm performance as they are the only ones of direct interest to powder based fusion. Alternative metrics for visual imaging such as the Structural Similarity Index (SSIM) would be of great benefit if implemented.
- **Completed translations** - Initial translations were completed with the help of Dr Ralf Mouthaan, Ms Sophia Gruber, and Mr Daoming Dong. These were not kept upto date with the program evolution and can be considered to be placeholder only.

- **Parameter hierarchy version updaters** - While simple changes to the parameter hierarchy does not break save files, significant changes such as moving parameters does. Newtonsoft provides a set of tools that can easily be used to convert between different JSON parameter file versions.
- **Masking regions** - HoloGen currently includes a basic tool set for drawing and editing regions of interest in an image. The work flow is limited though and the possibility to load the region data from an independant image file is planned. Using pixel intensity to choose the region of interest is also be planned.
- **Staged algorithms** - The design of the algorithms used has allowed for the possibility of an algorithm pipeline. Instead of pre-defining algorithms, users could drag and drop individual components such as FFTs into a list and Cuda's dynamic compilation could be used to generate the final algorithm type.
- **Expanded test suite** - HoloGen's code and UI tests at the moment are limited. A set of production quality tests is desired.

Appendix H

ResinPrinter Application Architecture

Appendix H

ResinPrinter Application Architecture

H.1 Introduction

Section 6.2 presented the resin demonstrator but didn't discuss the software in detail. Unlike HoloGen, the ResinPrinter application is relatively straight forward drawing its core functionality from HoloGen. Section H.2 summarises the key libraries, Section H.3 covers interfaces of note and Section H.4 explains the server architecture.

H.2 Libraries

- *ResinPrinter.App* - Top level library for the application itself. Handles persistence, settings, console commands and user interaction.
- *ResinPrinter.Manager* - Concrete implementation of the abstract interface defined in *ResinPrinter.Interfaces*. Responsible for managing the interactions and timings of the other modules.
- *ResinPrinter.Generator* - Concrete implementation of the abstract interface defined in *ResinPrinter.Interfaces*. Responsible for hologram generation.
- *ResinPrinter.Monitor* - Concrete implementation of the abstract interface defined in *ResinPrinter.Interfaces*. Responsible for monitoring of the other modules.

- *ResinPrinter.Aligner* - Concrete implementation of the abstract interface defined in *ResinPrinter.Printer*. Responsible for alignment of the camera, projectors and stage.
- *ResinPrinter.Slicer* - Concrete implementation of the abstract interface defined in *ResinPrinter.Interfaces*. Responsible for slicing of triangle meshes.
- *ResinPrinter.Photogrammetry* - Responsible for analysing the recorded images
- *ResinPrinter.Native.Wrapper* - Managed C++ code that wraps *ResinPrinter.Native.Cuda* and *ResinPrinter.Native.Slicer* within a .NET managed framework.
- *ResinPrinter.Native.Cuda* - C++ library implementing all code that operates within the Cuda framework.
- *ResinPrinter.Native.Slicer* - C++ library implementing slicing code for triangle meshes.
- *ResinPrinter.Native.Base* - C++ library providing base utilities for *ResinPrinter.Native.Cuda* and *ResinPrinter.Native.Slicer*. - Defines abstract interfaces for the concrete implementations in *ResinPrinter.Manager*, *ResinPrinter.Generator*, *ResinPrinter.Monitor*, *ResinPrinter.Aligner*, *ResinPrinter.Slicer*, etc.

Each individual library is detailed in the manual pages for that library.

H.3 Important Interfaces

- `ICam` - Generic camera service model API.
- `IBaslerCam` - Service model API for Dart series cameras manufactured by Basler. Extends `ICam`.
- `IWebCam` - Service model API for Microsoft Direct Show (DS) interfaced cameras. Extends `ICam`.
- `IThorCam` - Service model API for all ThorLabs cameras. Extends `ICam`.
- `IMiricleCam` - Service model API for Thermoteknix 307k Miricle cameras connected over USB and RS232. Extends `ICam`.
- `ILaserController` - Service model API for ILX Lightwave Laser Diode Controllers connected over GPIB.

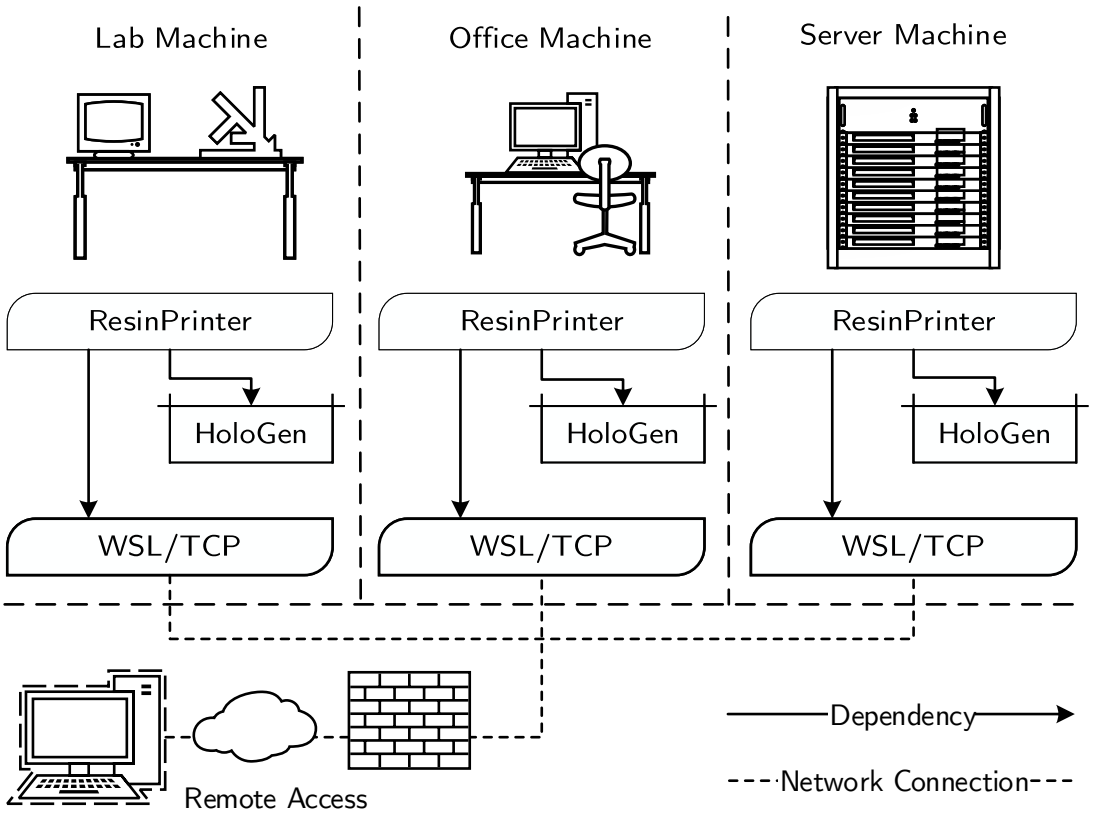


Fig. H.1 ResinPrinter server architecture.

- `ISPILaser` - Service model API for SPI fibre lasers connected over RS232.
- `IFunctionGenerator` - Service model API for function generator.
- `IPowerMeter` - Service model API for ThorLabs power meters.
- `IRotationMount` - Service model API for ThorLabs K-Cube Controllers
- `IStage1Axis` - Service model API for a single axis Melles Griot stage.
- `IStage3Axis` - Service model API for a three axis Melles Griot stage.
- `ITemperatureSensor` - Service model API for ThorLabs temperature and humidity sensors.

H.4 Server Architecture

ResinPrinter is built on top of a client server architecture using the Windows Web Services (WWS) API Service Model. This allows for the layout in Figure H.1 where the experiment can be conducted on the lab machine, the hologram generation on a server and the two controlled by an office machine or remote.

Appendix I

Additional Resources

Appendix I

Additional Resources

A number of companion resources are available.

I.1 HoloGen

Installer: www.gitlab.com/CMMPEOpenAccess/HoloGen/tree/master/Publish
Source code: www.gitlab.com/CMMPEOpenAccess/HoloGen
Test and benchmarking suite: www.gitlab.com/CMMPEOpenAccess/HoloGenTest
Initial specification: www.gitlab.com/PublicResearchData/HoloGenSpecification
Application manual: www.gitlab.com/PublicResearchData/HoloGenManual
Source code documentation: www.gitlab.com/PublicResearchData/HoloGenDocumentation

I.2 Experiment Resources

Risk assessments: https://gitlab.com/PublicResearchData/PhDRiskAssessments
Custom component CAD files: https://gitlab.com/PublicResearchData/PhDCustomComponents
Setup CAD assembly files: https://gitlab.com/PublicResearchData/PhDExperimentalSetups
Risk Assessments: https://gitlab.com/PublicResearchData/PhDRiskAssessments
LIDT testing animation: www.youtube.com/watch?v=0ayICJvRfME

Benchmarking raw data: www.gitlab.com/PublicResearchData/HoloGenBenchmarking
Jasper JD8714 phase drift: https://gitlab.com/PublicResearchData/JasperPhaseDrift/

I.3 Lab Resources

Melles Griot stage controller[1]: www.gitlab.com/CMMPEOpenAccess/melles-griot-stage-controller
Managed code interop demo: www.gitlab.com/CMMPEOpenAccess/Cuda-Cpp-CS-VB-Interop
CGH using Matlab: https://gitlab.com/CMMPEOpenAccess/CGH-Matlab-Demos
CGH using Python: https://gitlab.com/CMMPEOpenAccess/CGH-Python-Demos
DirectShow camera acquisition: https://gitlab.com/CMMPEOpenAccess/DirectShowCameraAcquisition

I.4 CAD Models

Thermoteknix Miricle 307K: www.gitlab.com/CMMPEOpenAccess/ThermoteknixMiricle307K-CAD
Goniometer: www.gitlab.com/CMMPEOpenAccess/GoniometerCAD
Thermo Scientific Digital Dry Bath: www.gitlab.com/CMMPEOpenAccess/digitaldrybath
Jasper SLM: https://gitlab.com/CMMPEOpenAccess/JasperSLMModel
Jasper SLM driver: www.gitlab.com/CMMPEOpenAccess/jasperdriverbox
Santec SLM-300: www.gitlab.com/CMMPEOpenAccess/SantecSLM-300
Coherent-Surelock-LM-Series: www.gitlab.com/CMMPEOpenAccess/Coherent-Surelock-LM-Series

I.5 Extended Edition

In addition to the above, the extended version of this thesis can be obtained from the author on request.

[1] George S. D. Gordon deserves credit for the initial reverse engineering of the communications syntax.